

# Sistema de Ayuda a la Toma de Decisiones basado en Árboles de Decisión Fuzzy: Estudio de un caso

---

Proyecto Fin de Máster en Ingeniería Informática para la Industria

Máster en Investigación en Informática, Facultad de Informática,  
Universidad Complutense de Madrid

Autor: Javier Rivas Rodríguez

Director: Matilde Santos, Departamento de Arquitectura de  
Computadores y Automática (DACYA), UCM

Colaborador de dirección: M<sup>a</sup> Carmen García-Alegre, Instituto de  
Automática Industrial (IAI), CSIC

Curso académico: 2008/09



El abajo firmante, matriculado en el Master de Investigación en Informática de la Facultad de Informática, autoriza a la Universidad Complutense de Madrid (UCM) a difundir y utilizar con fines académicos, no comerciales y mencionando expresamente a su autor el presente trabajo Fin de Master: “Sistema de Ayuda a la Toma de Decisiones basado en Árboles de Decisión Fuzzy: Estudio de un caso”, realizado durante el curso académico 2008-2009 bajo la dirección de Matilde Santos y con la colaboración externa de la dirección de M<sup>a</sup> Carmen García-Alegre en el departamento de Arquitectura de Computadores y Automática (DACyA), y a la biblioteca de la UCM a depositarlo en el Archivo Institucional E-Prints Complutense con el objeto de incrementar la difusión, uso e impacto del trabajo en Internet y garantizar su preservación y acceso a largo plazo.

Javier Rivas Rodríguez

|           |  |           |
|-----------|--|-----------|
| <b>1</b>  | <b>RESUMEN</b>   | <b>6</b>  |
| <b>2</b>  | <b>INTRODUCCIÓN</b>  | <b>7</b>  |
| 2.1       | MOTIVACIÓN   | 7         |
| 2.2       | MARCO DEL TRABAJO: PROYECTO CENIT HESPERIA                       | 7         |
| 2.3       | SISTEMA DE AYUDA A LA TOMA DE DECISIONES (SATD)                  | 8         |
| <b>3</b>  | <b>SISTEMAS EXPERTOS</b>   | <b>10</b> |
| <b>4</b>  | <b>LÓGICA Y TEORÍA DE CONJUNTOS BORROSOS (FUZZY)</b>             | <b>12</b> |
| 4.1       | LOGICA BORROSA   | 12        |
| 4.2       | CONTROL FUZZY  | 13        |
| <b>5</b>  | <b>ARBOLES DE DECISIÓN BORROSA</b>                               | <b>15</b> |
| <b>6</b>  | <b>DESCRIPCIÓN DEL PROBLEMA</b>                                  | <b>19</b> |
| <b>7</b>  | <b>SOLUCIÓN PROPUESTA</b>  | <b>23</b> |
| 7.1       | ADQUISICIÓN DE CONOCIMIENTO                                      | 23        |
| 7.2       | ORGANIZACIÓN DEL CONOCIMIENTO                                    | 25        |
| 7.3       | DEFINICIÓN DE LAS VARIABLES EN TÉRMINOS FUZZY                    | 28        |
| 7.4       | REPRESENTACIÓN DEL CONOCIMIENTO MEDIANTE ÁRBOL DE DECISIÓN FUZZY | 30        |
| 7.4.1     | INFERENCIA DE LA DECISIÓN  | 31        |
| 7.5       | SISTEMA DE AYUDA A LA TOMA DE DECISIONES (SATD)                  | 37        |
| 7.5.1     | MÓDULO DE ADQUISICIÓN DE DATOS                                   | 39        |
| 7.5.2     | MÓDULO DE PREPROCESO Y FUZZYFICACIÓN                             | 41        |
| 7.5.3     | MÓDULO DE RAZONAMIENTO Y TOMA DE DECISIONES                      | 42        |
| 7.5.4     | MÓDULO DE SALIDA   | 42        |
| <b>8</b>  | <b>RESULTADOS</b>  | <b>44</b> |
| <b>9</b>  | <b>COMENTARIOS, CONCLUSIONES Y TRABAJO FUTURO</b>                | <b>49</b> |
| <b>10</b> | <b>REFERENCIAS</b>   | <b>51</b> |
| <b>11</b> | <b>APENDICE A: DISEÑO DEL SATD</b>                               | <b>54</b> |
| <b>12</b> | <b>APÉNDICE B: IMPLEMENTACIÓN</b>                                | <b>58</b> |
| 12.1      | PAQUETE ENTRADA  | 58        |

|      |                               |    |
|------|-------------------------------|----|
| 12.2 | PAQUETE DE FUZZY              | 59 |
| 12.3 | PAQUETE RAZONAMIENTO          | 62 |
| 12.4 | PAQUETE DE LÓGICA DE PROGRAMA | 62 |

|    |  |           |
|----|--|-----------|
| 13 | <u>APENDICE C: ARTICULO PRESENTADO PARA JORNADAS DE AUTOMÁTICA,<br/>TARRAGONA 2008</u> | <u>64</u> |
|----|--|-----------|

## 1 Resumen

En este trabajo se presenta un sistema de razonamiento cualitativo que utiliza como soporte un árbol de decisión borroso para la vigilancia del nivel de agua de los embalses y la generación de alertas tempranas en situaciones previas a un posible desbordamiento. El objetivo es detectar situaciones anormales para prevenir con antelación a la población circundante. La aplicación desarrollada tiene como entradas, valores relativos al estado del embalse, de los ríos que confluyen en el mismo y de las condiciones meteorológicas. El razonamiento experto se modela mediante un árbol de decisión borroso, cuyas variables son las etiquetas lingüísticas que maneja el operario humano en su tarea de vigilancia y control.

This work presents a qualitative reasoning system, which uses a fuzzy decision tree as reasoning model, to monitor the water level of dams, and to generate warnings, previous to a possible overflow. The aim is to detect previously abnormal situations to warn the surrounding population. The developed application receives as input the values related to the dam state, the rivers which meet in it, and the weather. The expert reasoning is modelled by a fuzzy decision tree, whose variables are the linguistic labels with which the dam worker deals, to make his vigilance and control tasks.

Keywords: fuzzy control, fuzzy decision tree, expert system decision tree, fuzzy reasoning, approximate reasoning, soft computing, dam, security system.

Palabras clave: control borroso, árbol de decisión borroso, sistema experto, árbol de decisión, razonamiento borroso, computación suave, presa, sistema de seguridad.

## 2 Introducción

### 2.1 Motivación

Existe una necesidad común de proteger y garantizar la seguridad de infraestructuras de grandes dimensiones como edificios gubernamentales, aeropuertos o estaciones de energía. Esta demanda es especialmente fuerte en los territorios de la Unión Europea y en Estados Unidos donde las situaciones de crisis desencadenan desestabilizaciones en cadena que pueden llegar a afectar en gran medida a multitud de sectores de la sociedad.

Una crisis es un evento generalmente súbito y con escaso tiempo de reacción en donde las decisiones pueden estar afectadas por una gran cantidad de variables que conllevan muchos elementos de incertidumbre. Por ello, la utilización de las técnicas contenidas en las llamadas tecnologías de la información puede facilitar enormemente la gestión de los datos para la toma de decisiones adecuadas que conduzcan a paliar la crisis con los recursos disponibles, en el menor tiempo posible y con el menor número de pérdidas.

Existe un gran vacío en la aplicación de tecnologías en los procesos de gestión de crisis. Así, disponer de un volumen organizado de información y usar sistemas de gestión de la misma será crucial en todas las etapas, permitiendo optimizar la planificación y la toma de decisiones. Cuando, como en una situación de crisis, la información distribuida espacialmente es decisiva en la decisión, el sistema de gestión de la información deberá incluir algún mecanismo para gestionar adecuadamente la información espacial y las relaciones espaciales entre los elementos que integran la crisis y la gestión de la misma.

A las dificultades inherentes para tratar la información asociada a los sistemas de representación espacial, de audio y vídeo cognitivo, hay que sumar los problemas habituales de las grandes aplicaciones como son la interfaz de usuario, la arquitectura, las redes de comunicación, la seguridad, etc. Por ello, para afrontar todos estos problemas es necesario un desarrollo conjunto en varias áreas de investigación capaz de integrar todos los requisitos necesarios para el buen funcionamiento del sistema final.

### 2.2 Marco del trabajo: Proyecto CENIT Hesperia

El proyecto *Hesperia* tiene por objeto el desarrollo de tecnologías que permitan la creación de sistemas punteros de seguridad, vídeo vigilancia y control de operaciones de infraestructuras y espacios públicos. El proyecto surge para dar respuesta a una demanda sostenida a medio y largo plazo, en particular, en países de la Unión Europea y en Estados Unidos. La gestión integrada de seguridad y control de operaciones permitirá la implantación de sistemas rentables que, en este momento, no existen en el mercado.

Las tecnologías del proyecto resolverán la seguridad en dos tipos de escenarios. Por un lado, permitirán gestionar la seguridad y las operaciones de infraestructuras públicas especialmente sensibles, como subestaciones eléctricas, en gas, depósitos de agua o estaciones de telecomunicaciones. Por otro, incrementarán de forma sustancial los niveles de seguridad de grandes espacios públicos, como aeropuertos, estaciones de ferrocarril, puertos, centros de ciudades especialmente en zonas peatonales, centros comerciales, etc.

El CDTI, organismo adscrito al Ministerio de Industria Turismo y Comercio, creó en 2005 un programa de Consorcios Estratégicos Nacionales en Investigación Técnica (*CENIT*), cuyo principal objetivo es fomentar la cooperación público privada en I+D+i mediante la financiación de proyectos conjuntos de investigación industrial. El organismo ha dotado a CENIT con doscientos millones de euros durante cuatro años, el 46,5 % de la inversión prevista para todo el programa. Los otros 230 millones de euros los aportará el sector privado.

El consorcio está integrado por Indra Software Labs, Unión Fenosa, Tecnobit, SAC Control, Technosafe, Visual Tools y Brainstorm Multimedia. Asimismo, participan las Universidades de Castilla La Mancha, de Granada, de Extremadura, la Politécnica de Madrid, la de las Palmas, la Politécnica de Valencia y la Politécnica de Cataluña. La lista se completa con la colaboración del Consejo Superior de Investigaciones Científicas (CSIC) y el Centro Tecnológico del País Vasco (Ikerlan)[].

### 2.3 Sistema de Ayuda a la Toma de Decisiones (SATD)

En el pensamiento científico tradicional la comprensión de un fenómeno depende de la capacidad de analizarlo en términos cuantitativos precisos. Sin embargo los esquemas de razonamiento humano sugieren mecanismos de inferencia aproximados en cuanto a la génesis de las unidades informáticas, conceptos, y al establecimiento de las relaciones entre ellos, reglas.

El cerebro aprovecha la tolerancia en precisión para codificar la información relevante mediante términos lingüístico que conllevan una relación aproximada con los datos cuantitativos, asociándolos a percepciones y generando acciones [1].

En este marco, Zadeh presenta las bases iniciales de la teoría de conjuntos borrosos [2] con gran auge hasta el momento actual, aplicándose a áreas tan variadas como son el control automático, la optimización, clasificación, reconocimiento de patrones, etc.

Por otro lado y dentro de los modelos de razonamiento, se encuentran los árboles de decisión caracterizados por su capacidad de dividir un proceso complejo de toma de decisiones, en una secuencia de decisiones simples encadenadas que proporcionan un esquema de razonamiento fácil de interpretar [3],[4].

La fusión de conjuntos borrosos (fuzzy) y árboles de decisión permite combinar el manejo de la incertidumbre y habilidades del razonamiento aproximado, con la sencillez de



comprensión y fácil aplicación que ofrecen los árboles de decisión, potenciando su aplicabilidad en contextos complejos con incertidumbre [5].

Los árboles de decisión fuzzy (FDT) asumen que todos los atributos del dominio o variables lingüísticas, se han definido previamente en forma de conjuntos borrosos [6]. Se diferencian de los árboles de decisión tradicionales en dos aspectos fundamentales:

- Utilizan un criterio de bifurcación basado en restricciones borrosas
- Los procedimientos de inferencia son diferentes.

La selección de los conjuntos borrosos que definen los términos lingüísticos que se usan para construir el árbol, se realiza en el algoritmo, siguiendo un determinado criterio [6]. Esto se debe a que en un modelo de representación mediante reglas borrosas, el valor numérico de una variable puede tener un grado de pertenencia distinto de cero, en varios conjuntos borrosos, y al no existir restricciones se consideran las contribuciones de cada una de las reglas activadas en el resultado final.

En este trabajo se presenta, un sistema de razonamiento soportado por árboles de decisión borrosa como ayuda a la vigilancia y control del nivel de agua en los embalses, con la intención de detectar situaciones pre-críticas y enviar alertas tempranas.

El trabajo se ha estructurado de la siguiente forma, los fundamentos teóricos en los que se basa corresponden a las primera secciones, la sección 2 resume la idea general que hay detrás de los sistemas expertos, la sección 3 trata la lógica y la teoría de conjuntos fuzzy y en la sección 4 se explica el fundamento y la idea que hay detrás de los árboles de decisión fuzzy. En las siguientes secciones se pasa al tratamiento más directo del caso que se trata en concreto, en la sección 5 se explica el dominio del problema que se pretende resolver, y en la sección 6 se pasa a su descripción. En la sección 7 se detalla la solución propuesta para la resolución del problema aplicando arboles de decisión fuzzy. En la sección 8 se muestran los resultados obtenidos con ejecuciones del sistema implementado y en la sección 9 se pasa a comentar las conclusiones y posibles líneas de trabajo futuro. En los apéndices del trabajo se baja de nivel de abstracción y en el Apéndice A se explica el diseño que se ha seguido para llevar a cabo el Sistema de Ayuda a la Toma de Decisiones (SATD), y en la B se explica con mayor detalle su implementación. Por último en el Apéndice C se adjunta el artículo presentado y aceptado en las Jornadas de Automática 2008 en Tarragona, en el que se fundamenta este trabajo.

### 3 Sistemas Expertos

Un sistema experto se puede definir como una aplicación que contiene el conocimiento de un experto en un determinado dominio, y lo pone a disponibilidad de usuarios no expertos [7]. El sistema utiliza el conocimiento sobre el dominio y una estrategia de control específica para llegar a las soluciones.

Entre los primeros sistemas expertos se encuentra, en 1965, el DENDRAL que determinaba la estructura molecular a partir de la masa espectrométrica; el R1 usado para la configuración de sistemas de ordenador y el MYCIN utilizado con fines de diagnóstico médico. Desde la mitad de los años 60 se crearon muchos sistemas expertos en diferentes campos desde operaciones espaciales, hasta sistemas de decisión sobre pacientes en unidades de cuidados intensivos en hospitales [ 7].

Se califica al sistema experto como sistema y no como programa, porque aúna diferentes componentes, tales como una base de conocimiento, mecanismos de inferencia, base de hechos, etc. Todos estos componentes interactúan entre sí simulando el proceso de solución del problema que realiza un experto en el dominio [8].

Dado que un sistema experto pretende emular la actividad de los expertos humanos, debería disponer de las siguientes características:

- Resolución del problema que se plantea de manera análoga a como lo haría un experto humano
- Trabajo con datos incompletos o información imprecisa
- Capacidad de aprendizaje
- Reestructuración de los conocimientos de los que dispone en función de nuevos datos o información

En un sistema experto se consideran tres componentes básicos: la base de hechos, la base de conocimientos y el motor de inferencias [9], Figura 1.

La base de conocimiento, está formada por toda la información disponible sobre el dominio en el que se desarrolla la aplicación.

La base de hechos, está constituida por los datos sobre la situación concreta que va a ser considerada. Además puede utilizarse como memoria auxiliar o de trabajo para almacenar los modos de razonamiento. De esta forma el sistema es capaz de explicar como ha conseguido llegar al resultado final.

El motor de inferencia se encarga de gestionar la información existente en la base de conocimiento y los datos de la base de hechos, relacionándolos para razonar sobre ellos y obtener un resultado.

En una aplicación, se toman decisiones o se establecen hechos de manera secuencial. Este proceso de utilizar hechos actuales y conocimiento contenido en la base de conocimiento

para establecer hechos adicionales o decisiones, continua en forma de cadena hasta llegar a un hecho especificado como objetivo. El mecanismo de control lleva a cabo, principalmente, un procesamiento simbólico denominado inferencia. El conocimiento que contienen las reglas puede utilizarse de diferentes maneras en el proceso de inferencia. Por consiguiente, el mecanismo de control puede consistir en varias y diferentes estrategias de inferencia[8].

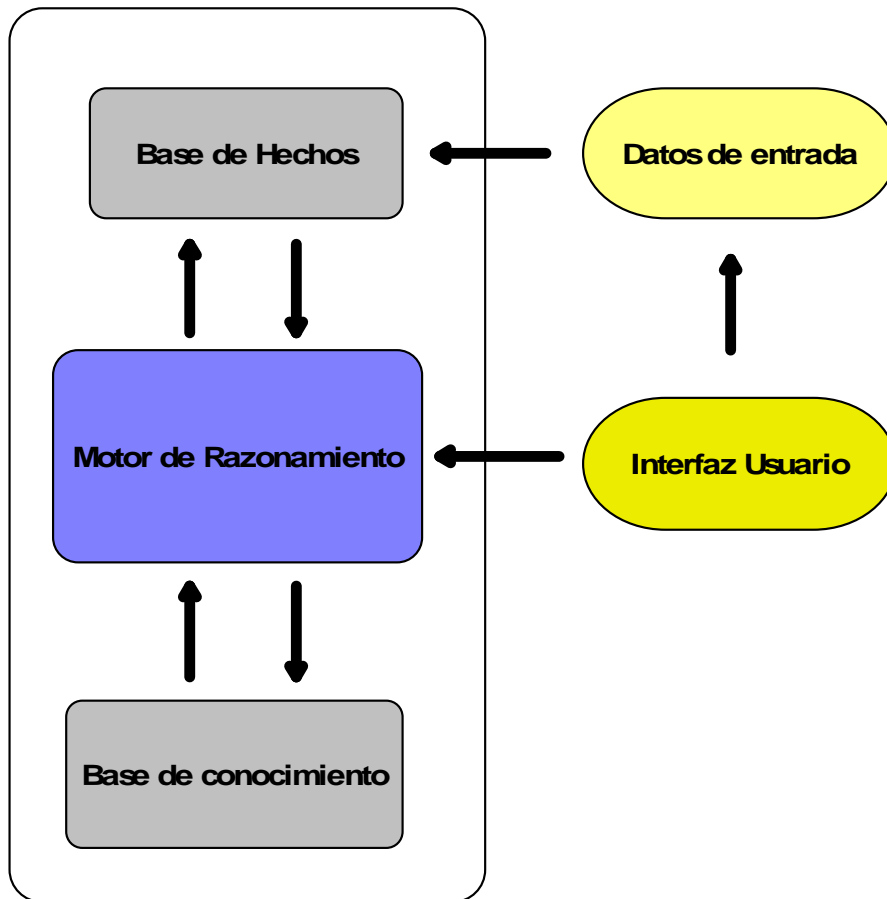


Figura1: Arquitectura de un Sistema Experto

Las ventajas que plantean estos sistemas automáticos frente a los expertos humanos son varias. Por una parte el conocimiento contenido en los sistemas expertos es más fácil de documentar y de transferir que el de los expertos humanos. Por otra parte ese conocimiento es remanente, es decir, permanece tras la desaparición de los expertos. Asimismo resulta fácilmente transportable y no está sujeto a factores que pueden afectar al rendimiento humano como cansancio, enfermedad o presión [9].

## 4 Lógica y Teoría de conjuntos borrosos (Fuzzy)

En los conjuntos tradicionales, un elemento del universo puede pertenecer o no a un conjunto. Es decir, la pertenencia de un elemento a un conjunto es nítida (“crisp”) [7].

La teoría de conjuntos borrosos se basa en el reconocimiento de que determinados conjuntos poseen unos límites imprecisos. Estos conjuntos están constituidos por colecciones de objetos para los cuales la transición de pertenecer o no pertenecer a los mismos no es abrupta sino gradual. Así, un conjunto borroso se caracteriza por una función de pertenencia ( $\mu_A(t)$ ) definida como un número real en el intervalo  $[0,1]$ . En casos extremos, si el valor de la función es 0, el elemento no pertenece al conjunto, y si es 1, el elemento pertenece al 100% al conjunto.  $\mu_A(t)$  indica el grado de compatibilidad del valor asignado a la variable  $t$ , con el concepto representado por  $A$ . Donde  $A$  es el valor lingüístico (concepto, etiqueta lingüística) asociado al subconjunto borroso definido por  $\mu_A(t)$  [10],[11].

Para que se comprenda mejor el concepto y utilidad de los conjuntos fuzzy, se muestra el siguiente ejemplo basado en [11]:

Sean las personas de una organización cualquiera el universo de discurso. Un subconjunto de todos los hombres en esta organización será un subconjunto crisp. Ahora bien, considerando el conjunto de la gente “joven”, obviamente la “juventud” no es una función escalón con valores 0 o 1 a una cierta edad, los que poseen edad inferior a 30 años; de forma natural, se asociaría un grado de juventud a cada elemento, en función de su edad, por ejemplo, {persona1 valor 0.8, persona2 valor 0.1, persona3 valor 1}. Donde, la persona2 podría tener 28 años, la persona2, 50 años y la persona3, 23 años.

La utilidad de un conjunto borroso para modelar un determinado concepto o etiqueta lingüística depende de lo apropiada que sea la definición de su función de pertenencia. Esto es de suma importancia en todas aquellas situaciones en las que se manejan términos del lenguaje natural.

### 4.1 Lógica borrosa

La lógica estudia métodos y principios de razonamiento en todas las formas posibles. La lógica clásica maneja proposiciones que pueden ser verdaderas o falsas. Las primeras lógicas n-evaluadas fueron propuestas por Lukasiewicz en 1930 como extensión de su lógica Trievaliada (verdadero, falso e indeterminado) [10].

Teniendo en cuenta el isomorfismo que existe entre la lógica clásica y la Teoría de Conjuntos, se establece otro isomorfismo entre la lógica estándar n valuada de Luka L1 y

la teoría de conjuntos borrosos original. Ambas son a su vez isomorfas con un Algebra de Boole. Ver Tabla1.

| Teoría de Conjuntos | Algebra de Boole | Lógica proposicional |
|---------------------|------------------|----------------------|
| $P(X)$              | $B$              | $F(v)$               |
| $\cup$              | $+$              | $\cup$ disyunción    |
| $\cap$              | $\cdot$          | $\wedge$ conjunción  |
| $-$                 | $-$              | $-$ negación         |
| $X$                 | $1$              | $1$ verdadero        |
| $\emptyset$         | $\emptyset$      | $\emptyset$ falso    |
| $\subseteq$         | $\leq$           | $\leq$ desigualdad   |

Tabla 1: Correspondencia entre lógica de conjuntos, Algebra de Boole y Lógica Proposicional [10]

$P(X)$  representa el conjunto formado por todos los subconjunto del conjunto  $X$ ,  $B$  el conjunto que tiene al menos dos elementos  $(0,1)$  y  $F(v)$  el conjunto de todas las proposiciones que son verdaderas.

Estos isomorfismos van a permitir abarcar todas las teorías desarrollando únicamente una de ellas. Como los conjuntos fuzzy son una extensión de los conjuntos clásicos, la lógica fuzzy es una extensión de la lógica clásica, por tanto de la misma manera que se cumple el isomorfismo entre la lógica clásica y la teoría de conjuntos clásica, existe por tanto también entre la teoría de conjuntos fuzzy y la lógica fuzzy. Ver tabla2 [11]

| Teoría de conjuntos Fuzzy | Lógica Fuzzy  |
|---------------------------|---|
| $\mu_A(x)$                | valores de verdad de la proposición "x es miembro del conjunto A" |
| $\cap$                    | $\wedge$ conjunción   |
| $\cup$                    | $\cup$ disyunción   |
| $\neg$                    | $-$ negación  |

Tabla 2: Correspondencia entre teoría de conjuntos fuzzy y lógica fuzzy [11]

La lógica borrosa constituye una extensión de las lógicas multievaluadas. Su objetivo último es dotar de fundamentos al razonamiento aproximado con proposiciones imprecisas. Utilizando la teoría de conjuntos borrosos como herramienta principal. El elemento primario de la lógica borrosa es el lenguaje natural, donde usualmente se utiliza el razonamiento aproximado con proposiciones imprecisas [10].

## 4.2 Control fuzzy

Los sistemas de control fuzzy basados en reglas se pueden considerar un subconjunto de los sistemas expertos. A su vez, difieren de los sistemas clásicos de control en su intento de modelar fundamentalmente la habilidad de un operador en lugar de modelar el proceso en sí.

Su utilización permite afrontar situaciones en las que se presentan:

- Sistemas complejos
- Sistemas parcialmente definidos, o definidos de manera imprecisa.
- Variables difíciles de medir, o incluso de determinar
- Sistemas sujetos a grandes perturbaciones del entorno o a perturbaciones debidas a variaciones inevitables en la calidad del material utilizado.

La aproximación de un sistema de control basado en reglas difiere de la aproximación convencional en el uso de información cualitativa en lugar de relaciones analíticas precisas para controlar un determinado proceso. De esta forma se simula la categorización de la información que realizan los operadores expertos al controlar un proceso mediante etiquetas lingüísticas [10].

En un sistema de reglas fuzzy se tienen reglas del tipo:

Si (X es ETIQUETA1) AND (Y es ETIQUETA2) entonces (Z es ETIQUETA3)

Donde X e Y son variables de entrada y Z es la variable de salida. Dados unos valores específicos de X e Y, la tarea consiste en determinar el valor de Z utilizando las reglas de control aplicables y la inferencia fuzzy.

Se definen variables fuzzy que pueden representar los valores de las variables de entrada y de salida. Típicamente los valores lingüísticos que puede alcanzar una variable pueden ser: Negativo Alto, Negativo Medio, Negativo Bajo, Cero, Positivo Bajo, Positivo Medio, Positivo Alto.

El siguiente paso sería definir las funciones de pertenencia para las variables fuzzy. El tipo de función de pertenencia (trapezoidal, triangular, Gaussiana, etc) que se elija depende de multitud de factores, como el tipo de aplicación, la precisión del control requerido, etc.

Posteriormente se definirían las reglas que relacionan la combinación de variables mediante los operadores de agregación, definidos en la lógica fuzzy, con la salida del sistema.

Por último se procede al cálculo de la acción neta, que se realiza mediante una combinación adecuada de las acciones individuales que derivan de cada regla, pesadas con el grado de verdad calculado para las condiciones del antecedente de cada una de ellas.

## 5 Árboles de decisión borrosa

Los árboles de decisión son una de las opciones más utilizadas para el razonamiento y el aprendizaje basado en ejemplos, para tratar con el lenguaje natural y la incertidumbre. Han sufrido una serie de alteraciones a lo largo del tiempo a fin de mejorar su capacidad de representación e inferencia [12].

En este caso, se trata de una modificación que permita combinar las decisiones simbólicas, con el razonamiento aproximado que ofrece la representación del universo fuzzy. Se intenta explotar las ventajas que ofrecen ambas visiones: la amplia utilización del aprendizaje desde ejemplos, la facilidad de comprensión de los árboles de decisión, y la habilidad para manejar la incertidumbre de la información de la representación fuzzy.

La fusión de ambas técnicas utiliza metodologías ya existentes en ambas áreas, para obtener la máxima ventaja de ellas. En particular, las inferencias de conocimiento deben ser definidas nuevamente para los árboles de decisión borrosa (fuzzy), en este caso basadas en sistemas de reglas y control fuzzy.

El conocimiento resultante, en forma de árbol de decisión y procedimientos de inferencia, ha sido elogiado por su facilidad de comprensión [12]. Esto apela a un amplio rango de usuarios que están interesados en el dominio? de la comprensión de las capacidades de clasificación o de la simbología de las reglas que podrían extraerse de un árbol [13] y por consiguiente ser aplicadas en un sistema basado en reglas. Este interés, en cambio, ha generado un gran esfuerzo en investigación teniendo como resultado grandes avances tanto metodológicos como empíricos [14], [15], [16].

Los árboles de decisión se popularizaron a través del trabajo de Quinlan [17] con el algoritmo de clasificación ID3. Los sistemas que se basaron en esta aproximación han dado buenos resultados en dominios simbólicos. Los árboles de decisión asignan decisiones simbólicas a nuevos casos, lo que hace que sean difíciles de aplicar en aquellas situaciones en las que se necesita una decisión numérica, o cuando la decisión numérica mejora el procesamiento subsiguiente. Por otra parte según Quinlan [18], los resultados de los árboles de decisión son categóricos y por lo tanto no tienen en cuenta incertidumbres potenciales en la clasificación. Así, basta con que haya pequeños cambios en los valores de los atributos en un caso que este siendo clasificado, para que puedan dar lugar a cambios inapropiados en la asignación de la clase. Como solución Quinlan propone en [18] un método probabilístico de construcción de árboles de decisión como clasificador. Sin embargo, el tipo de incertidumbre que surge en los problemas de clasificación no tiene porque ser necesariamente probabilístico, apareciendo con aleatoriedad y ruido [19].

Hay varios tipos de incertidumbre, que pueden clasificarse en dos grandes categorías: estadísticas y cognitivas.

La incertidumbre estadística trata con información que aparece debido a un comportamiento aleatorio de los sistemas físicos. La incertidumbre cognitiva de naturaleza muy diferente a la estadística, trata con fenómenos que provienen del pensamiento, razonamiento y percepción humanos.

Se puede refinar aun más la clasificación de la incertidumbre cognitiva en dos subcategorías: vaguedad y ambigüedad [19].

En general, el concepto de vaguedad se asocia con la dificultad de marcar un límite o definir distinciones en el mundo, de la misma manera que un domino de interés se considera vago si no se puede delimitar claramente por un umbral. Por otra parte, la ambigüedad se asocia a una relación de uno a muchos, como pueden ser situaciones con dos o más alternativas de tal manera que la elección entre ellos se deja sin especificar [20]

La aproximación fuzzy es una de las posibles extensiones. La lógica de conjuntos fuzzy proporciona una base para la representación, permitiendo modelar la incertidumbre relativa al lenguaje, mientras que por otra parte proporciona un marco que facilita la comprensión del conocimiento. [19].

Las técnicas en las que se aplican los árboles de decisión, se han mostrado hasta ahora interpretables, eficientes, independientes del problema y capaces de tratar con aplicaciones a gran escala, pero también han sido reconocidas como clasificadores altamente inestables con respecto perturbaciones mínimas con los datos de entrenamiento, es decir, son métodos que presentan una alta variabilidad [21]. La lógica fuzzy supone una mejora en estos aspectos, debido a la flexibilidad del formalismo de los conjuntos fuzzy.

Estudios detallados de estos métodos comparados alternativamente con métodos crisp han demostrado una gran mejora en la precisión de la predicción, debido principalmente a la reducción del modelo de variabilidad [21]. También, los parámetros presentan mayor estabilidad (casi 50% menos de variabilidad que en los árboles de decisión clásicos) y esto hace que se pueda llegar a una mejor interpretación [22].

En sistemas de reglas fuzzy, las reglas proporcionan una gran facilidad de comprensión y transfieren el conocimiento de alto nivel, mientras que la teoría de conjuntos, junto con la lógica fuzzy y los métodos de razonamiento aproximado proporcionan la habilidad de modelar los detalles del conocimiento [12]. Por consiguiente, la representación fuzzy está aumentando su popularidad para la resolución de problemas con incertidumbre, ruido o falta de precisión en los datos [22], siendo aplicada de manera satisfactoria a problemas en el dominio industrial[12].

El objetivo por tanto es fusionar las habilidades de manejo, con el razonamiento aproximado de la representación fuzzy y de los árboles de decisión, conservando las ventajas de ambos. Por un aparte el manejo de la incertidumbre y el procesamiento gradual del primero, con la facilidad de comprensión, popularidad y facilidad de aplicación de los segundos.



Esto aumentará el poder de representación y por consiguiente la capacidad de aplicación de los árboles de decisión, añadiendo una componente adicional de conocimiento basada en el paradigma fuzzy [12]. Se espera que estas modificaciones, que demuestran que las ideas y técnicas de la Inteligencia Artificial y los sistemas fuzzy pueden ser fusionados, proporcionen árboles de decisión fuzzy más inmunes a la imprecisión, aumentando su aplicación en contextos de incertidumbre. Al mismo tiempo, debería aumentar la facilidad de comprensión, manteniendo la misma estructura arbórea como componente principal del resultado del razonamiento [23].

Los árboles de decisión fuzzy se diferencian de los árboles de decisión tradicionales en dos aspectos fundamentales:

- Utilizan criterios de ramificación basados en restricciones fuzzy
- Los procesos de inferencia son diferentes. Los conjuntos fuzzy, definen los criterios (fuzzy), impuestos en el algoritmo, que se usa para la construcción del árbol.

A continuación se define un método de aplicación de algoritmos de árboles de decisión fuzzy general, basado en la propuesta de [12].

Para este trabajo se utilizan únicamente conjuntos fuzzy representados mediante funciones trapezoidales. Normalmente el árbol no está podado, y cada nodo interno tiene una ramificación, por cada valor lingüístico de la variable consultada, salvo cuando los casos de entrenamiento no satisfacen alguna restricción. Para solucionar este problema, se asume que durante el proceso de inferencia se asigna valor desconocido a un nodo, cuando el valor que se necesita no aparece en el caso que se está tratando, o cuando el nodo no tiene ramificación para él.

La notación utilizada para este caso, es la siguiente:

- El conjunto de las variables fuzzy es  $V = \{V_1, V_2, \dots, V_n\}$   
para cada variable  $V_i \in V$ :
  - o Datos crisp son  $u^i$  pertenece a  $U_i$
  - o  $D_i$  es el conjunto de etiquetas lingüísticas fuzzy de cada variable.
  - o  $V_p^i$  denota el valor de la etiqueta lingüística  $p$  de la variable  $V_i$ .
- El conjunto de etiquetas lingüísticas para la variable decisión es  $D_c$
- Para cada nodo  $N$  del árbol de decisión fuzzy:
  - o  $F^N$  es el conjunto de etiquetas lingüísticas, de los nodos ascendentes en el recorrido hasta el nodo  $N$ .
  - o  $V^N$  es el conjunto de variables que se consultan en el recorrido hasta  $N$ .
  - o  $N \mid v_p^i$  es el hijo en concreto de  $N$  que se crea a partir de la ramificación en función de la variable  $V_i$ , que cumple  $v_p^i$  "pertenece"  $D_i$ .
  - o  $X_N = \{X_j^N\}$  es el conjunto de la computación de las pertenencias de los casos de ejemplo  $E$  para el nodo  $N$ .
  - o  $S_N^V$  es el conjunto de hijos del nodo  $N$ , utilizando  $V_i \in \{V - V^N\}$  como variable de ramificación.

Procedimiento de construcción del árbol de decisión fuzzy:

- Se comienza procesando los ejemplos a través del nodo raíz del árbol
- Para cualquier nodo  $N$  que no se haya expandido se calculan las pertenencias de los ejemplos
- Para cada nodo, se busca el conjunto de variables que faltan por ser computadas  $V - V^N$  para ramificar el nodo
- Se ramifica el nodo  $N$  en  $D_i$  subnodos. El nodo hijo  $N \mid v_{p^i}$  hereda los ejemplos definidos por  $X^{N \mid v_{p^i}}$ . Las nuevas funciones de pertenencia se calculan usando las etiquetas lingüísticas,  $V_{p^i}$  que recorren el árbol hasta el nodo  $N$ .

## 6 Descripción del problema

El escenario a analizar es un embalse o presa genérica, con aportes de uno o varios ríos. Una presa se define como una barrera fabricada con piedra, hormigón o materiales sueltos, que se construye habitualmente en una cerrada o desfiladero sobre un río, arroyo o canal con la finalidad de retener el agua en el cauce fluvial para: posterior aprovechamiento en abastecimiento o regadío, elevar su nivel con el objetivo de derivarla a canalizaciones de riego, o para la producción de energía mecánica al transformar la energía potencial del almacenamiento en energía cinética, y ésta nuevamente en mecánica al accionar, mediante la fuerza del agua un elemento móvil. La energía mecánica puede aprovecharse directamente, como en los antiguos molinos, o de forma indirecta para producir energía eléctrica, como se hace en las centrales hidroeléctricas [24].

La elección de este dominio viene dado por formar parte de uno de los escenarios estratégicos designado como potencialmente peligrosos, sensible a la provocación de situaciones de emergencia en caso de grandes avenidas.

Una situación de alerta que desemboque en una mala gestión del riesgo, por parte del personal a cargo puede provocar daños tales como, ruptura del muro de contención de la presa, inundaciones en el entorno, agravándose en el caso de existir poblaciones cercanas, etc. En este caso, las situaciones de alerta las pueden provocar: avisos de bomba, incendio en los mecanismos de la presa, o desbordamiento del agua embalsada. En situaciones de deshielo y lluvias el nivel del agua aumenta considerablemente con riesgo de superar el muro del dique de contención, Figura 2, [25].

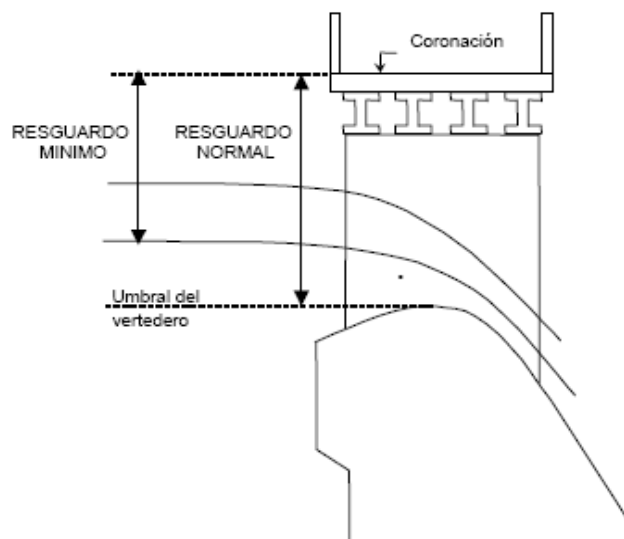


Figura 2: Sección transversal del muro de un embalse

A continuación se definen algunos términos del vocabulario del dominio, de interés para el diseño del sistema de razonamiento propuesto [26]:

- Coronación: Superficie superior que delimita el embalse
- Aliviadero: Estructura hidráulica por la que rebosa el agua excedente cuando el embalse está lleno.
- Cota : Marca el nivel de ocupación del embalse, se mide en metros
- Resguardo: Diferencia de cota entre el nivel de agua del embalse y el nivel de coronación
- Aforo: Aportaciones de agua que recibe el embalse. Estas aportaciones se componen principalmente de la lluvia y del caudal del río, que puede o no ser controlado en función de la existencia de otros embalses aguas arriba. El aforo se mide en m<sup>3</sup>/s.
- Avenida: Elevación del nivel de un curso de agua muy superior a su flujo medio, y que puede dar lugar a vertidos por coronación
- Turbinado: El volumen de agua desalojado por el embalse, que pasa a través de las turbinas, generando así, la energía que produce la central hidroeléctrica.
- Vertido: Volumen de agua desalojado por el embalse a través de las compuertas y/o el aliviadero. Al no pasar a través de las turbinas, no genera energía. El vertido se provoca cuando se prevé riesgo de coronación por avenida.

En un embalse se llega a una situación de crisis extrema cuando el agua sobrepasa la altura del embalse y se producen vertidos por coronación. Como consecuencia, pueden verse dañados elementos estructurales, tales como las barandillas o el alumbrado, e incluso, en situaciones extremas y dependiendo del tipo de construcción del muro de contención, podrían aparecer fisuras, la ola de coronación puede causar inundaciones en los terrenos colindantes y en la población civil de los pueblos aguas abajo, por lo que es necesario prevenir estas situaciones mediante la detección del nivel de agua del embalse y las condiciones meteorológicas y medioambientales del momento de observación [27].

El escenario de aplicación es la central Hidroeléctrica de Belesar. El embalse de Belesar es un pantano artificial creado en 1963 en el río Miño (el más grande de este río). Situado entre los concellos de Taboada, Chantada, O Saviñao, Paradelá, O Páramo, Guntín y Portomarín. Su capacidad es de 654,5 hm<sup>3</sup>, ocupando una superficie de 2.000 hectáreas y presentando una cola de 50 km [30].



Figura 3: Imagen aérea de la presa de Belesar (Orense)



Figura 4: Imagen frontal del dique de la presa de Belesar (Orensa), con una cota muy baja.

El objetivo es diseñar un sistema de razonamiento automático aproximado, basado en el conocimiento experto para el control de la apertura de las compuertas y la determinación de un nivel de riesgo, en tiempo real.

Se trata de un sistema altamente no lineal, y en el que las variables que intervienen pueden tener un nivel bajo de precisión. La finalidad del prototipo es evitar la situación de crisis provocada por alguno de los agentes o la combinación de varios que pueden intervenir en el desbordamiento de la presa, detectando situaciones de pre-crisis.

El sistema de razonamiento debe llevar a cabo las tareas de:

- Vigilancia del estado del nivel de agua del embalse o presa
- Generación de alertas tempranas: riesgos previsibles, protocolos a activar y actuación sobre las compuertas en situaciones de mayor riesgo.

La generación de alertas tempranas es fundamental para preservar tanto las instalaciones de la presa como las de la población asentada en las proximidades del mismo.

El sistema debe tomar la misma decisión de actuación sobre las compuertas de la presa que tomaría el personal responsable de la vigilancia y actuación en la presa.

## 7 Solución propuesta

En este apartado se describe el proceso que se ha llevado a cabo para diseñar un sistema de razonamiento aproximado, con las características requeridas.

### 7.1 Adquisición de conocimiento

En este apartado se describe el proceso que se ha seguido para extraer el conocimiento del dominio.

Para que el sistema pueda llegar a tomar las decisiones correctas, es necesario establecer una etapa de adquisición de conocimiento, que permita obtener el máximo conocimiento que el experto pueda proporcionar acerca del dominio.

Con este fin se han realizado un conjunto de entrevistas con las personas responsables de las oficinas de la central Hidroeléctrica de Belesar en Orense, así como con los responsables de explotación de Unión Fenosa en Madrid. En estas reuniones se realizó, en primera instancia, un análisis detallado de las variables que intervienen en la toma de decisiones que realizan los operarios, y la interrelación de las mismas va a determinar el nivel de emergencia.

Las variables que afectan al nivel de agua de la presa son de índole atmosférico-medioambiental y están muy interrelacionadas, Figura 5.

Mediante el análisis de las variables, se llega a la conclusión de que las mismas siguen una estructura jerárquica en lo que respecta a las prioridades, en función de cómo afectan éstas, en última instancia, a la evolución del volumen de agua del embalse.

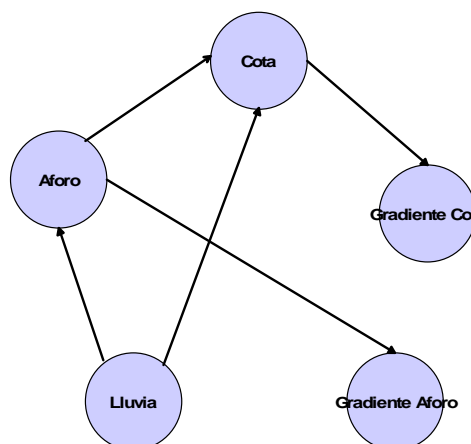


Figura 5: Diagrama de influencia de las variables seleccionadas. La punta de la flecha indica sobre qué variable se ejerce la influencia.

Las variables seleccionadas para el proceso de razonamiento son las siguientes:

- **Cota:** Esta variable, como se ha definido en secciones anteriores, mide la altura a la que llega el agua en el muro de contención del embalse. Por su condición, es la variable más crítica, ya que bajo ningún concepto el agua debe superar la marca de cota máxima. Se ve influenciada de manera directa por diferentes factores como pueden ser, las precipitaciones de la zona, crecidas del río aguas arriba, deshielo...etc. La medida de la misma se realiza mediante un sensor denominado cotímetro, que indica en todo momento el nivel del agua.
- **Lluvia:** Volumen de lluvia (mm/h) que afecta a la zona geográfica en la que se sitúa el embalse. La lluvia incide tanto en el caudal del río como en el nivel del agua del embalse. Para que este valor sea realmente significativo es necesario tomar las medidas de lluvia en un punto lo más próximo posible al embalse. Para medir ésta variable se usan pluviómetros repartidos a lo largo de la cuenca del río, y estaciones meteorológicas que proporcionan las medidas cada hora.
- **Aforo:** Indica el flujo (en  $\text{m}^3/\text{s}$ ) del río que desemboca en el embalse. El caudal del río se ve influenciado por la lluvia, y afecta directamente al nivel de agua del embalse.
- **Gradiente de cota:** Variación de la cota por unidad de tiempo. Es significativa, y de su análisis temporal se puede prever con mayor certeza, la proximidad a una situación de emergencia. La unidad de tiempo utilizada para el gradiente es 3 horas.
- **Gradiente de lluvia:** Variación de la lluvia cada 3 horas.

También se consideran otras variables que no dependen directamente del estado actual del sistema, que se calculan a partir del modelo, experimental o histórico, de las variaciones del nivel del embalse, lluvia y caudal en años anteriores.

La variable definida como histórico resulta de valorar la influencia de diferentes variables relacionadas con el contexto temporal, como son: la fecha actual, la estación del año, las condiciones meteorológicas y otros condicionantes de riesgo.

El histórico permite comparar el estado del sistema en años anteriores con respecto al estado actual, para detectar posibles anomalías o variaciones que ayuden a detectar situaciones de pre-emergencia. El tratamiento del histórico se contempla como una medida de prevención y contingencia de una situación de riesgo que no se ha dado todavía pero que, en caso de suceder, se desea evitar de inmediato.

Para la obtención de los históricos se han utilizado los datos almacenados en las bases de datos de la central hidroeléctrica de Belesar.



## 7.2 Organización del conocimiento

Una vez que se han determinado las variables que intervienen en el proceso de toma de la decisión, es necesario analizar las mismas así como sus interrelaciones para determinar las prioridades.

Con el fin de ajustar con la máxima precisión el conocimiento integrado en una representación en árbol de decisión y verificar el funcionamiento del sistema, se ha realizado un análisis de la evolución de las variables seleccionadas para realizar el proceso de toma de decisión sobre la apertura de compuertas que realiza el operario ante una situación de crisis en la presa Belesar. Se ha seleccionado una ventana temporal (año hidrológico) en el que se presentaron varias avenidas como resultado de las crecidas del caudal del río aguas arriba de la presa.

Estos datos, a su vez, serán los ejemplos que el sistema dispondrá como entrada. De esta manera, se puede comprobar si el resultado que proporciona el SATD diseñado es adecuado, y se puede extrapolar a los datos actuales.

Las variables disponibles para el análisis son:

- **Cota del embalse**

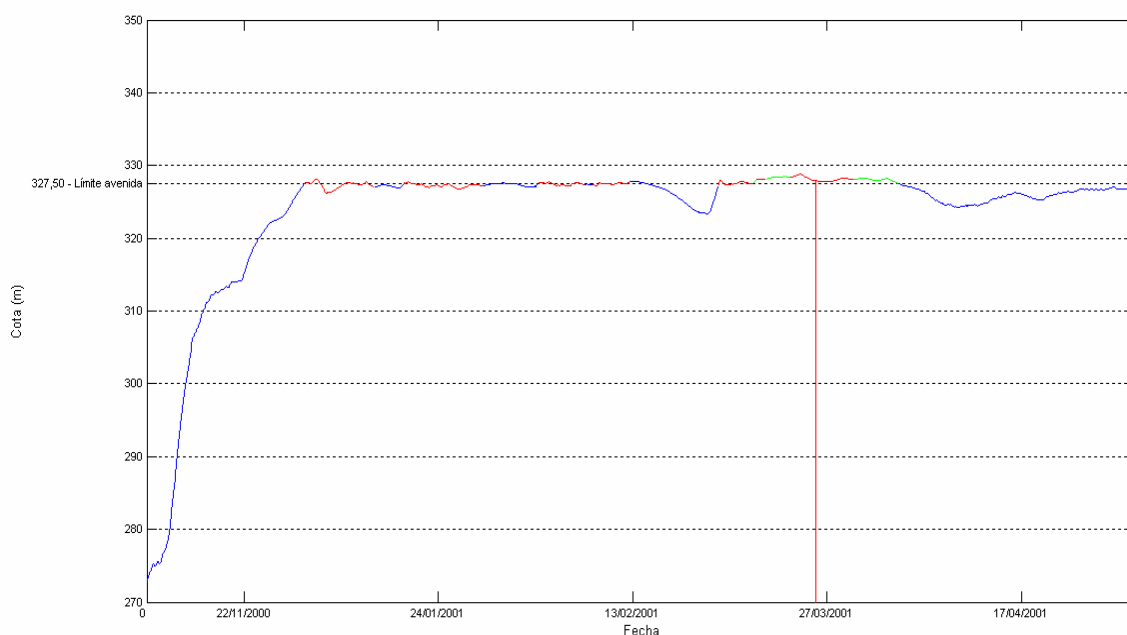


Figura 6: Datos horarios de la **Cota** de la presa de Belesar durante el periodo temporal comprendido entre el 01/11/2000, 00:00h y el 31/05/2001, 23:00h. Los tramos marcados en color rojo se corresponden con periodos en los que se dieron situaciones de **Avenida**, y los marcados en color verde se corresponden con periodos con perturbaciones en la recepción de datos.

- **Aforo**

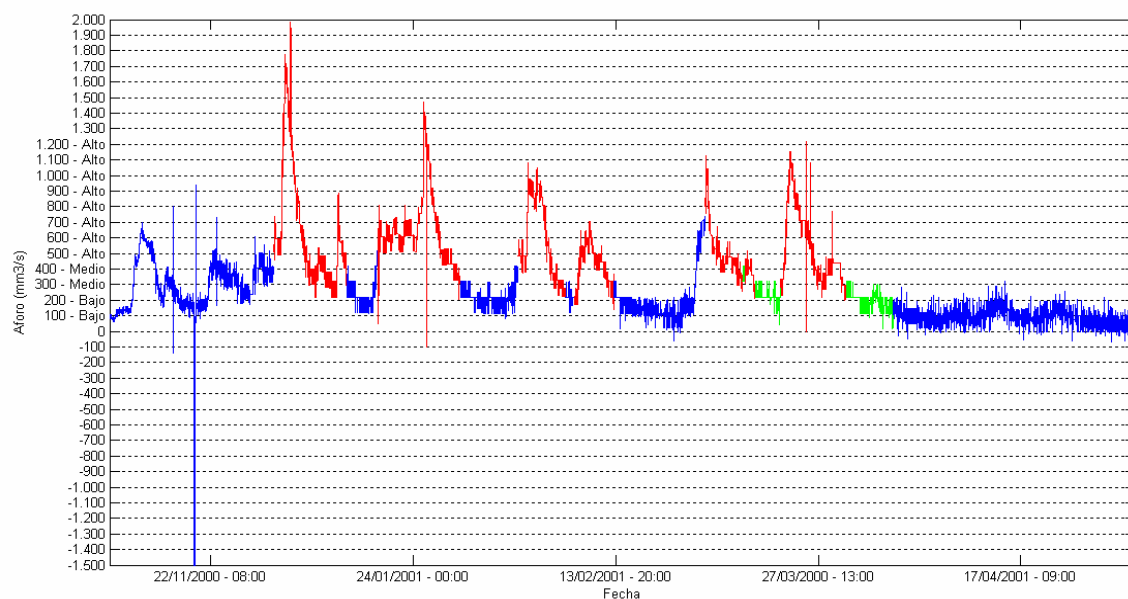


Figura 7: Datos horarios del **Aforo** de la presa de Belesar durante el periodo temporal comprendido entre el 01/11/2000, 00:00h y el 31/05/2001, 23:00h. Los tramos marcados en color rojo se corresponden con periodos en los que se dieron situaciones de **Avenida**, y los marcados en color verde se corresponden con periodos de perturbación en la recepción de datos

- **Volumen de agua turbinado**

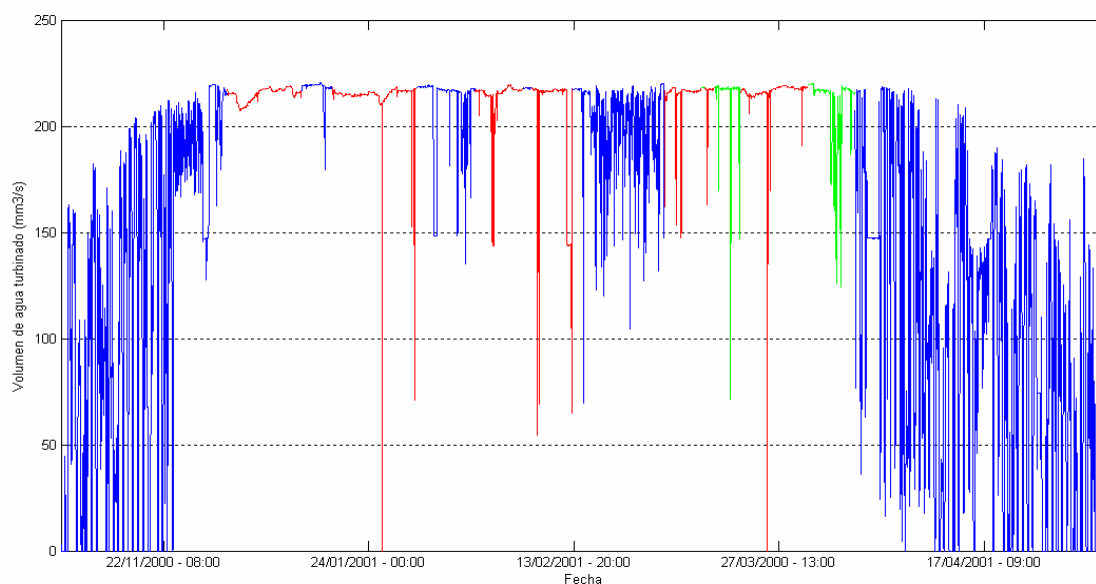


Figura 8: Datos horarios del volumen de agua turbinado en presa de Belesar durante el periodo temporal comprendido entre el 01/11/2000, 00:00h y el 31/05/2001, 23:00h. Los tramos marcados en color rojo se corresponden con periodos en los que se dieron situaciones de **Avenida**, y los marcados en color verde se corresponden con periodos de perturbación en la recepción de datos

- **Volumen de agua vertido**

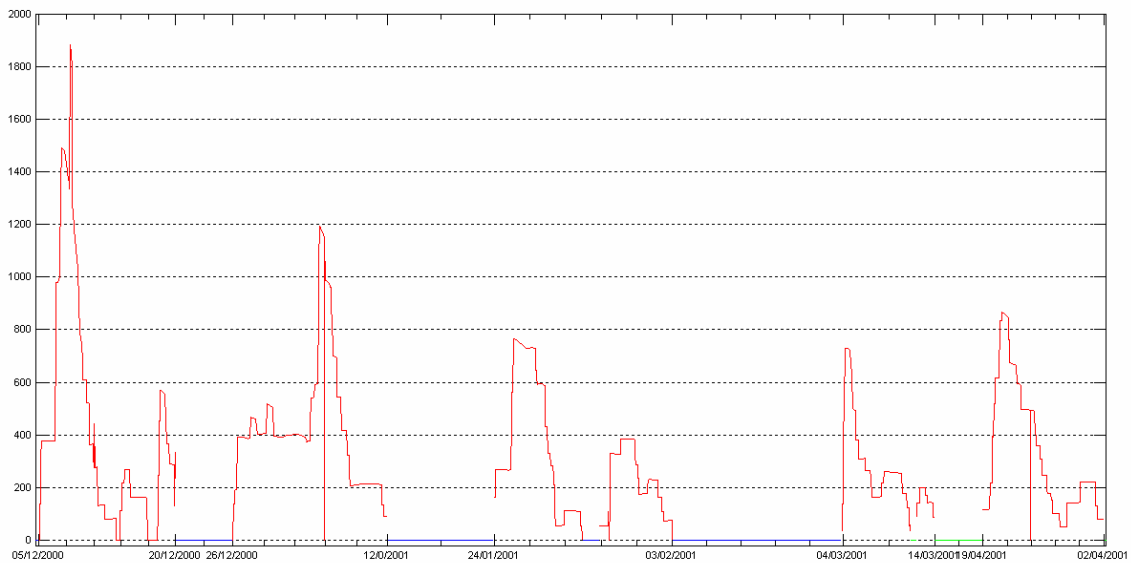


Figura 9: Datos horarios del **Volumen de agua vertido** en la presa de Belesar durante el periodo temporal comprendido entre el 01/11/2000, 00:00h y el 31/05/2001, 23:00h. En este caso el **Vertido** (apertura de todas las compuertas) solo se produce cuando hay **Avenida**

- **Volumen de lluvia**

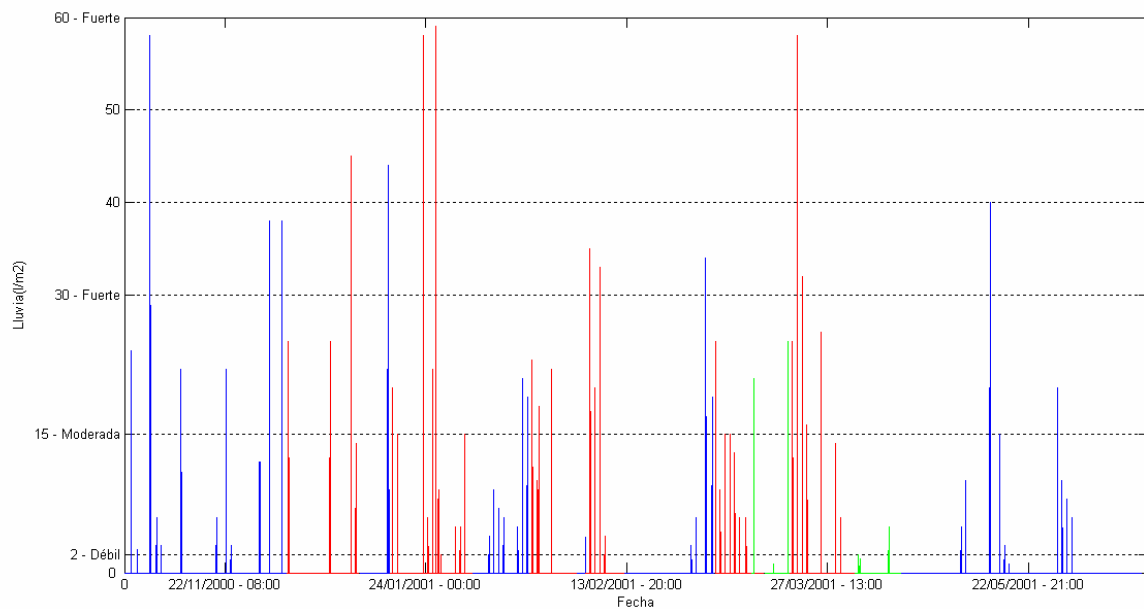


Figura 10: Datos horarios de la **Lluvia** en la presa de Belesar durante el periodo temporal comprendido entre el 01/11/2000, 00:00h y el 31/05/2001, 23:00h. Los tramos marcados en color rojo se corresponden con periodos en los que se dieron situaciones de **Avenida**, y los marcados en color verde se corresponden con periodos de fuerte anomalía en la recepción de datos

### 7.3 Definición de las variables en términos fuzzy

En este apartado se detallan los valores lingüísticos de las variables físicas seleccionadas y su representación mediante funciones de pertenencia trapezoidales. Los términos lingüísticos asignados a cada etiqueta, así como los límites de las funciones de pertenencia se han definido con la ayuda de expertos en el dominio, en concreto, la definición se llevó a cabo en las reuniones mantenidas con el personal de la central Hidroeléctrica de Belesar (Orense).

- **Cota**

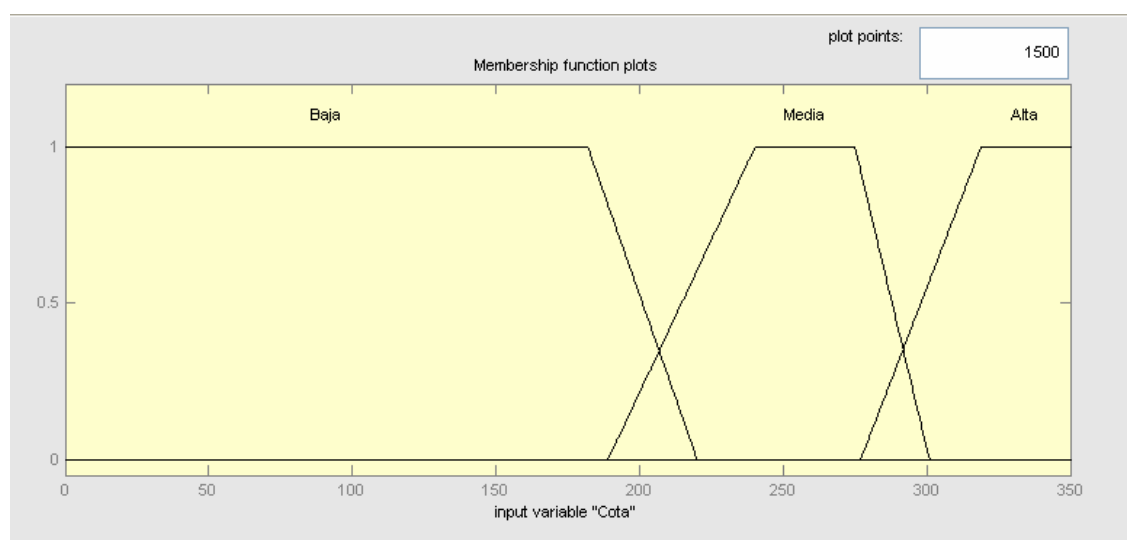


Figura 11: Variable Cota con tres valores lingüísticos {Baja, Media, Alta}.

- **Aforo**

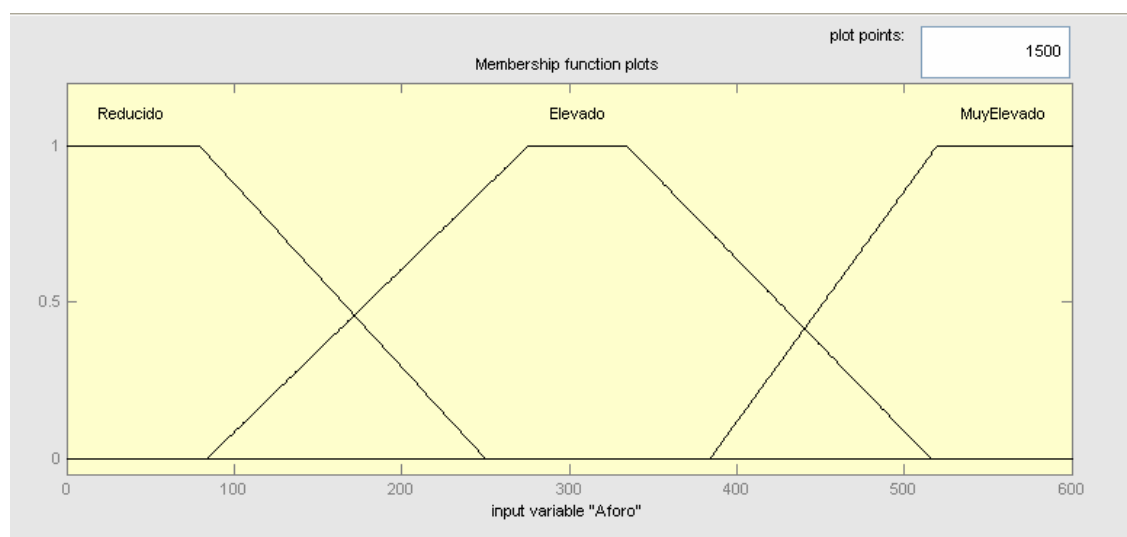


Figura 12: Variable Aforo con tres valores lingüísticos: {Reducido, Elevado y Muy\_Elevado}

- **Lluvia**

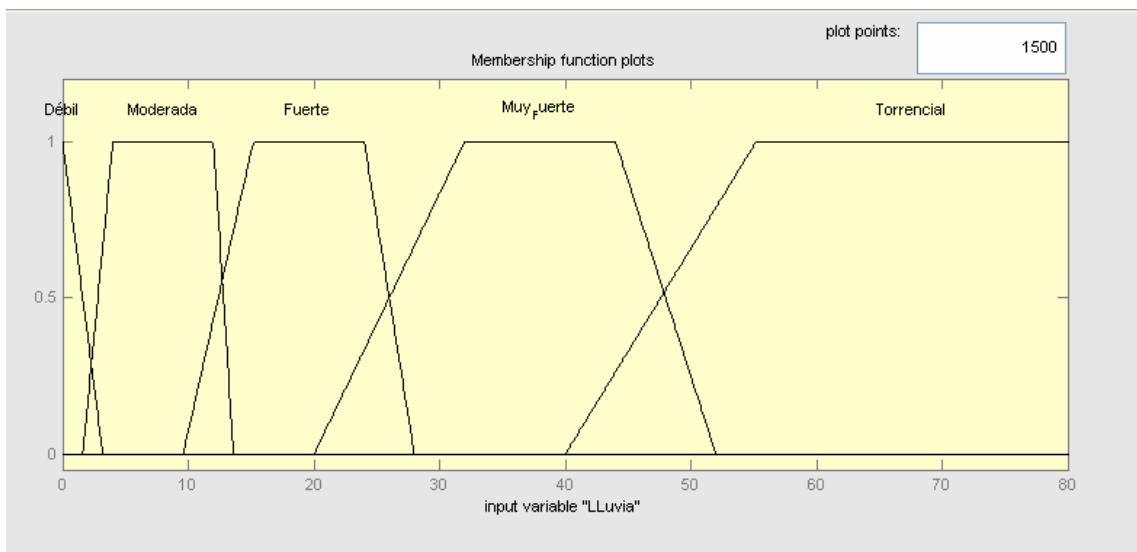


Figura 13: Variable Lluvia con cinco valores lingüísticos: {Débil, Moderada, Fuerte, Muy\_Fuerte y Torrencial.}

- **Gradiente Aforo**

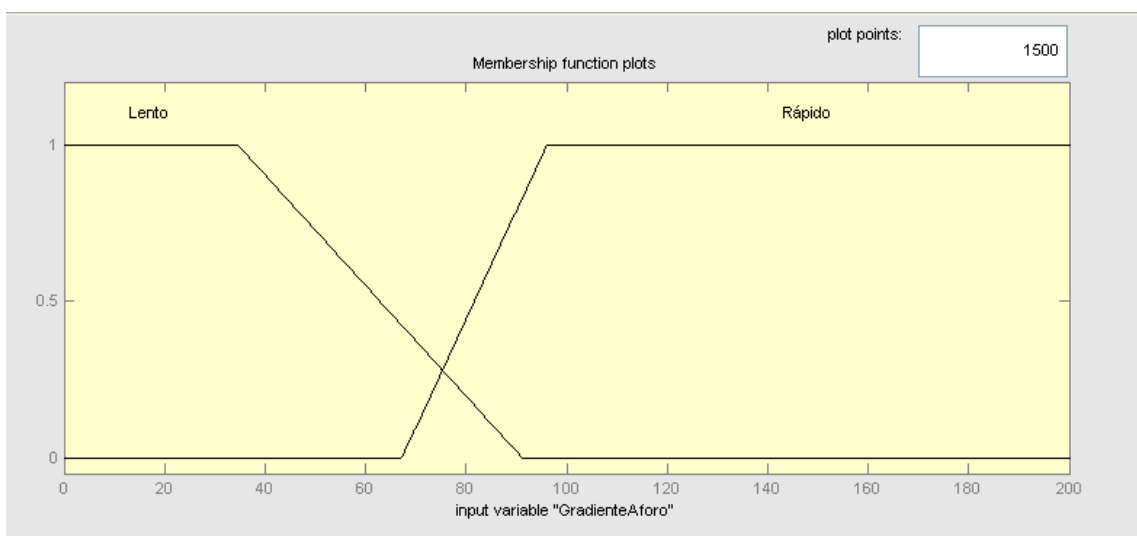


Figura 14: Variable Gradiente Aforo con dos valores lingüísticos {Lento y Rápido}

- **Gradiente Cota**

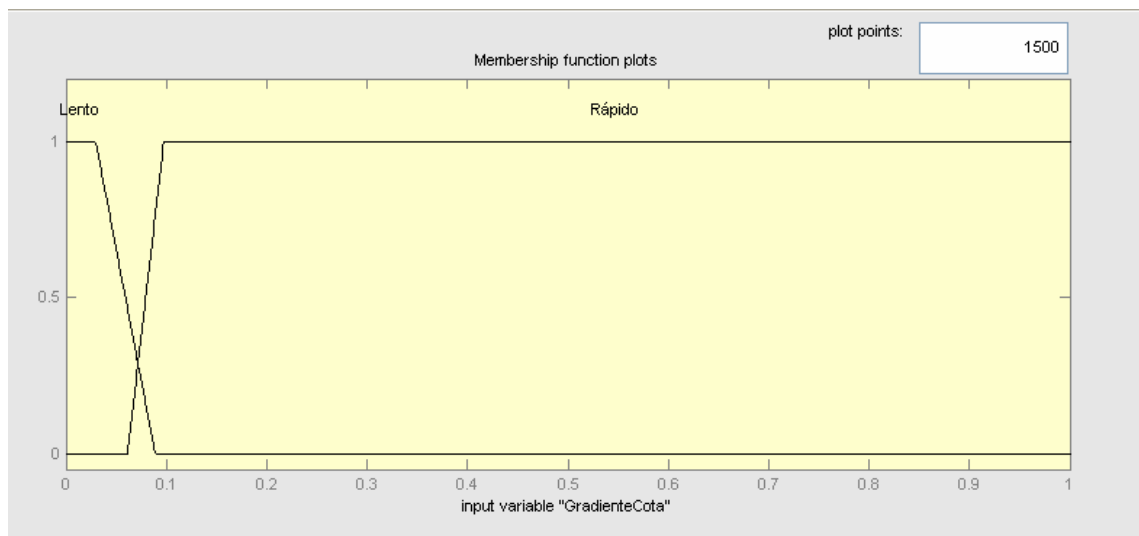


Figura 15: Variable Gradiente Aforo con dos valores lingüísticos {Lento y Rápido}

## 7.4 Representación del conocimiento mediante árbol de decisión fuzzy

En este apartado se muestra el modelo teórico de razonamiento mediante un árbol de decisión fuzzy.

Cómo se ha indicado en las secciones anteriores, con los árboles de decisión fuzzy, se pretende aunar, por una parte un manejo fácil y alta flexibilidad en el tratamiento de la incertidumbre proporcionada por la representación fuzzy de las variables, con la facilidad de comprensión y aplicación de los árboles de decisión.

Inicialmente se pensó en un modelo basado en un conjunto de reglas fuzzy, que integrasen el conocimiento experto y del dominio, del tipo:

**R:** IF (*entrada1 es  $vl_1$* ) AND (*entrada2 es  $vl_2$* ) AND (*entrada3 es  $vl_3$* )  
**THEN** (*salida1 es  $vl_{1m}$*  AND *salida2 es  $vl_{2n}$* )

Donde la codificación de las variables es como sigue:  $vl_i$  es el valor lingüístico  $j$  asignado a la entrada  $i$ .

El planteamiento de un sistema de razonamiento borroso con una Base de Conocimiento no constituye, en este caso, la mejor solución, por dos razones:

- 1) La combinación de las variables de entrada seleccionadas, con varias etiquetas lingüísticas cada una, dificulta al experto la generación de la Base de Conocimiento.

- 2) Las prioridades, bien determinadas, de unas variables sobre otras en los esquemas de razonamiento experto aconsejan otro modelo de representación del conocimiento.

Las dos consideraciones anteriores, llevan a la propuesta de un modelo de toma de decisiones jerárquico representado mediante un árbol de decisión borroso.

En el caso del problema de control al que se pretende dar solución, la relación entre las variables, viene marcada por una dependencia muy jerarquizada. De tal manera que es necesario utilizar algún mecanismo que ponga de manifiesto esta jerarquía. Pese a que la diferencia de relevancia entre las variables se puede resolver de diferentes formas, el esquema jerárquico que proporciona un árbol de decisión es el que más se ajusta a la realidad del problema que se busca modelar en este caso.

El dominio en el que se encuadra ésta aplicación, posee un número elevado de variables, por lo que es importante que la organización del conocimiento se lleve a cabo de una manera sencilla, intuitiva y fácil de estructurar para aproximar el método de razonamiento lo más posible al proceso que realiza el operario experto. Con un árbol de decisión, se plantea una posible solución a dicha organización, verificando las expectativas de sencillez y al mismo tiempo precisión en la representación del conocimiento.

Por otra parte, en este dominio, la decisión que toma el experto no viene unívocamente determinada en función de valores estrictamente numéricos. Utiliza su conocimiento y experiencia. Cuando las variables alcanzan determinados valores, la decisión que se toma no es predecible, dado que el grado de incertidumbre y la ambigüedad del umbral, que determina qué tipo de decisión que se ha de tomar, es alto. Al mismo tiempo, es necesario considerar, que los datos de entrada al sistema se obtienen mediante sensores de fluidos y la medida tiene una alta componente de ruido.

Ésta es una de las razones del modelo mediante conjuntos fuzzy, especialmente indicada para la resolución de problemas con incertidumbre, ruido o imprecisión en los datos [12].

Para el caso en concreto en estudio, no es posible el procesamiento simultáneo de todos los ejemplos de entrenamiento. Porque, como el objetivo del sistema es el tratamiento de la pre-crisis, y la toma de la mejor decisión para evitar una crisis, se ha de tomar decisión con cada ejemplo.

Debido a esto, el conjunto de ejemplos se procesa de manera individual y secuencial. Esto influye de manera decisiva en el proceso, tanto de construcción como de inferencia de la decisión, a partir del árbol de decisión fuzzy.

#### **7.4.1 Inferencia de la decisión**

Las inferencias de conocimiento se definen nuevamente para los árboles de decisión fuzzy, en este caso basadas en sistemas de reglas fuzzy. En el modelado que se presenta,

cada nodo del árbol se corresponde con una variable  $V_i$  “pertenece” a  $V$ , siendo  $V$  el conjunto de todas las variables.

Al existir una fuerte estructura jerárquica en las variables, aquellas que estén en los ordenes superiores del árbol tendrán mayor poder de discriminación y peso en la decisión, que aquellas que se encuentren en los niveles más alejados de la raíz. De esto se deduce que en el nodo raíz del árbol está la variable prioritaria y en el nivel del árbol en el que los nodos cuyos hijos sean las hojas, estará la variable de menor relevancia para la toma de decisión.

El proceso de inferencia es el siguiente:

- Se comienza explorando la raíz del árbol. Para ello se toma el valor crisp de entrada de la variable correspondiente.
- Se comprueba el valor de pertenencia  $\mu_{V_i^{D_i}}(e_i)$  a cada una de las etiquetas lingüísticas que definen la variable, siendo  $e_i$  el valor de la componente del ejemplo que se está cuantificando lingüísticamente, y  $D_i$  una etiqueta lingüística correspondiente a la variable  $V_i$ .
- De entre todos los valores de pertenencia, el algoritmo diseñado utiliza la función máximo para decidir qué etiqueta lingüística se ajusta al valor de entrada. De esta manera la variable ya está fuzzificada, y cuenta con un valor lingüístico, descartando el resto de posibles valores.
- Se ramifica el nodo variable raíz, y se elige el camino del árbol que lleva al siguiente nodo en función del valor fuzzificado del nodo padre.
- Este proceso se repite mientras queden variables por consultar, y no se haya llegado a un nodo hoja terminal.
- Una vez que se llega al final del árbol, el nodo hoja cuenta con la información relativa al recorrido que ha llevado hasta él y la decisión de actuación correspondiente.

La decisión relativa a las condiciones de entrada es un mensaje al operario que se encuentra en la sala de control de la presa, por lo que no hay que realizar un proceso de defuzzificación de la salida. El nodo hoja terminal ya contiene la información pertinente a la actuación a realizar.

A continuación se detalla, con un ejemplo, el proceso de razonamiento:

- Se está analizando el siguiente ejemplo(datos relativos a la presa Belesar 05/12/2000 9:00 A.M.)

| Cota   | Aforo  | Lluvia | Gradiente Cota | Gradiente Aforo | Histórico |
|--------|--------|--------|----------------|-----------------|-----------|
| 320,52 | 518,92 | 00,0   | 0,14           | 103,45          | Favorable |



- La variable de mayor relevancia es la **Cota**; por ello es la primera que se evalúa y se corresponde con la raíz del árbol. Se calcula la pertenencia del valor de la cota a los tres conjuntos fuzzy {Baja, Media, Alta}.
  - $\mu_{Cota}^{Baja}(327,52) = 0,0$
  - $\mu_{Cota}^{Media}(327,52) = 0,0$
  - $\mu_{Cota}^{Alta}(327,52) = 0,87$
- Aplicando la función máximo, resulta que el valor de la variable **Cota** es **Alta**, por tanto se expande el nodo Cota y se elige el camino correspondiente. Figura 16

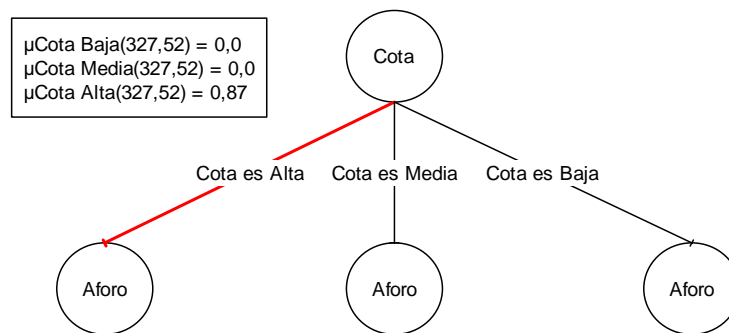


Figura 16

- Se continua evaluando la pertenencia de la siguiente variable, **Aforo**, en función del valor numérico de la variable:
  - $\mu_{Aforo}^{Reducido}(518,92) = 0,0$
  - $\mu_{Aforo}^{Elevado}(518,92) = 0,14$
  - $\mu_{Aforo}^{Muy\_Elevado}(518,92) = 0,88$
- Aplicando la función máximo, el valor de la variable **Aforo** es **Muy\_Elevado**, por tanto se expande el nodo Aforo y se elige el camino correspondiente. Figura 17

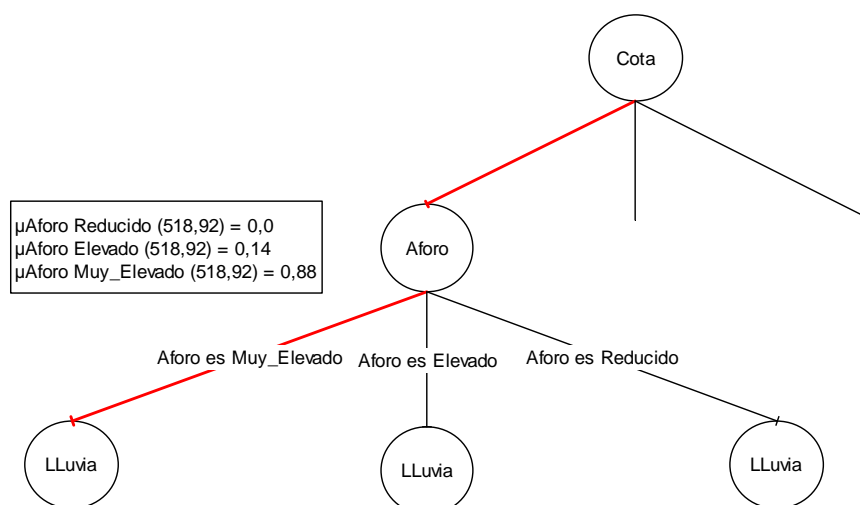


Figura 17

- Se continua evaluando la pertenencia de la siguiente variable, **LLuvia**, en función del su valor numérico:
  - $\mu_{LLuvia}^{Débil}(00,0) = 1,0$
  - $\mu_{LLuvia}^{Moderada}(00,0) = 0,0$
  - $\mu_{LLuvia}^{Fuerte}(00,0) = 0,0$
  - $\mu_{LLuvia}^{Muy\_Fuerte}(00,0) = 0,0$
  - $\mu_{LLuvia}^{Torrencial}(00,0) = 0,0$
- Aplicando la función máximo, resulta que el valor de la variable **LLuvia** es **Débil**, y se expande el nodo Lluvia y se elige el camino correspondiente. Figura 18

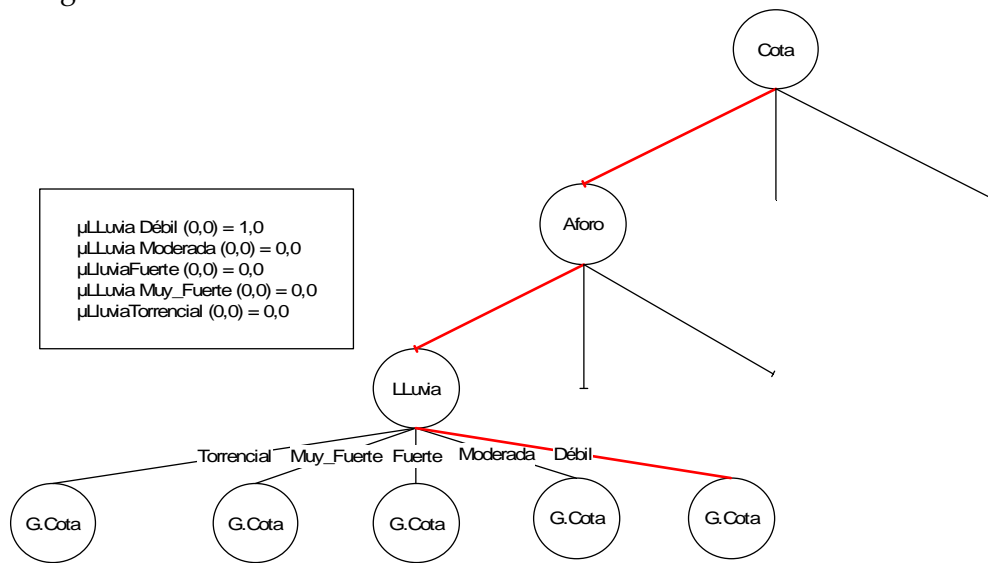


Figura 18

- Se continua evaluando la pertenencia de la siguiente variable, **Gradiente Cota**, en función de su valor numérico en la entrada de datos:
  - $\mu_{G.Cota}^{Lento}(0,14) = 0,0$
  - $\mu_{G.Cota}^{Rápido}(0,14) = 1,0$
- Aplicando la función máximo, resulta que el valor de la variable **Gradiente Cota** es **Rápido**. Se expande el nodo Gradiente Cota y se elige el camino correspondiente. Figura 19

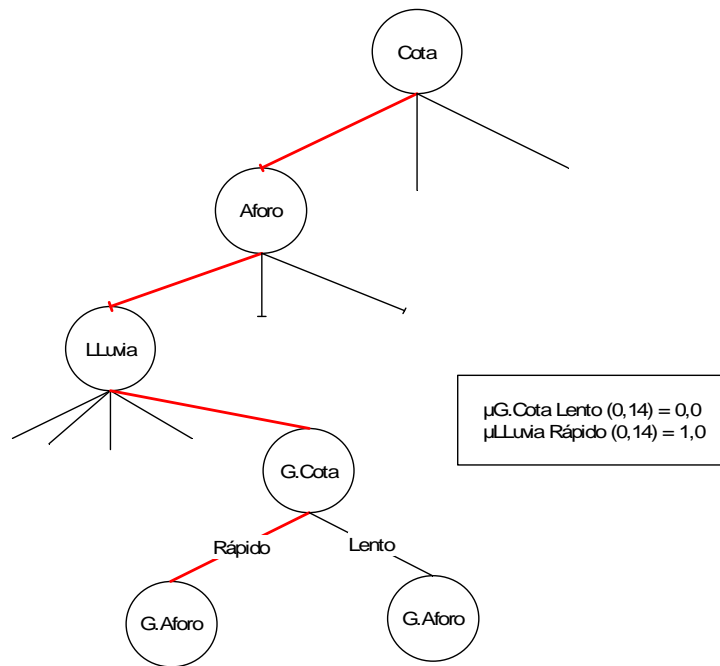


Figura 19

- Se evalúa la pertenencia de la siguiente variable, **Gradiente Aforo**, en función de su valor numérico:
  - $\mu_{G.Aforo \text{ Lento}}(103,45) = 0,0$
  - $\mu_{G.Aforo \text{ Rápido}}(103,45) = 1,0$
- Aplicando la función máximo, el valor de la variable **Gradiente Aforo** es **Rápido**, por tanto se expande el nodo Gradiente Aforo y se elige el camino correspondiente Figura 20

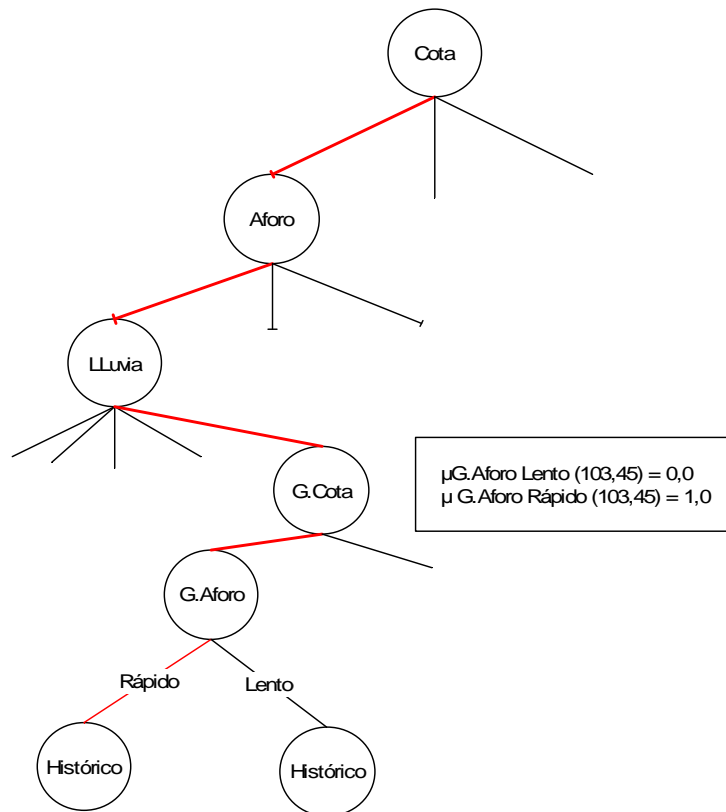


Figura 20

- En este caso ya no quedan más variables para evaluar, pero adicionalmente, se tiene en cuenta la variable Histórico. En éste caso no hay que calcular ninguna pertenencia; se expande el nodo hijo asociado al valor Histórico que se determina en la entrada. En este ejemplo de inferencia, el valor es **Favorable**. Con esto se indica, que las condiciones meteorológicas en situaciones anteriores han sido favorables, Figura 21.

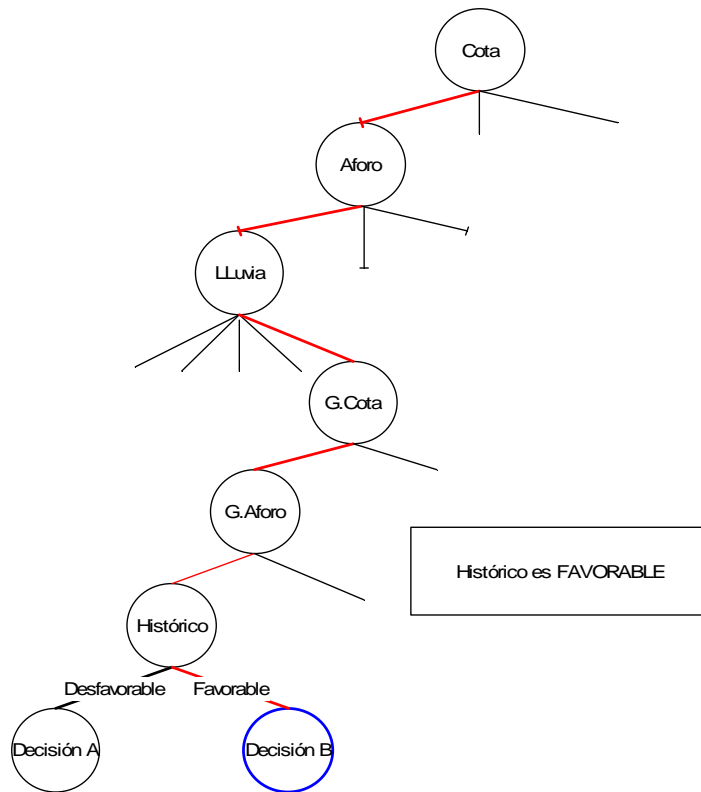


Figura 21

- Una vez se que se expande un nodo Terminal, la exploración del árbol finaliza, y el resultado de la toma de decisión es la información contenida en el nodo hoja terminal.

Con este ejemplo se muestra en detalle, el proceso de inferencia seguido para llevar a cabo la toma de decisión final. Los nodos terminales, contienen en última instancia la información sobre el tipo de actuación que se aconseja llevar a cabo con las compuertas de la presa. Finalmente, hacer notar que es la conjunción de todos los valores lingüísticos de las variables del antecedente de las reglas la que conduce al consecuente de la decisión.

## 7.5 Sistema de Ayuda a la Toma de Decisiones (SATD)

Se ha diseñado una aplicación, con interfaz gráfico para la visualización de estados y alertas, que se utilizará como demostrador para mostrar los resultados de la operación de diferentes algoritmos de razonamiento en la toma de decisiones. El objetivo del mismo es su implantación en el Centro de Control Integrado (CCI) de la Central Hidroeléctrica de Velle, que vigila las presas de Unión Fenosa, situadas en el río Miño. En este apartado se presenta, en alto nivel, el diseño del programa, su estructura en bloques y el flujo de información.

El sistema está pensado para funcionar de manera autónoma, con la mínima interrelación con el usuario final.

El sistema diseñado, se puede considerar en un primer nivel de grano grueso, Figura 22, como una caja negra que se relaciona con el exterior mediante dos entradas significativas.

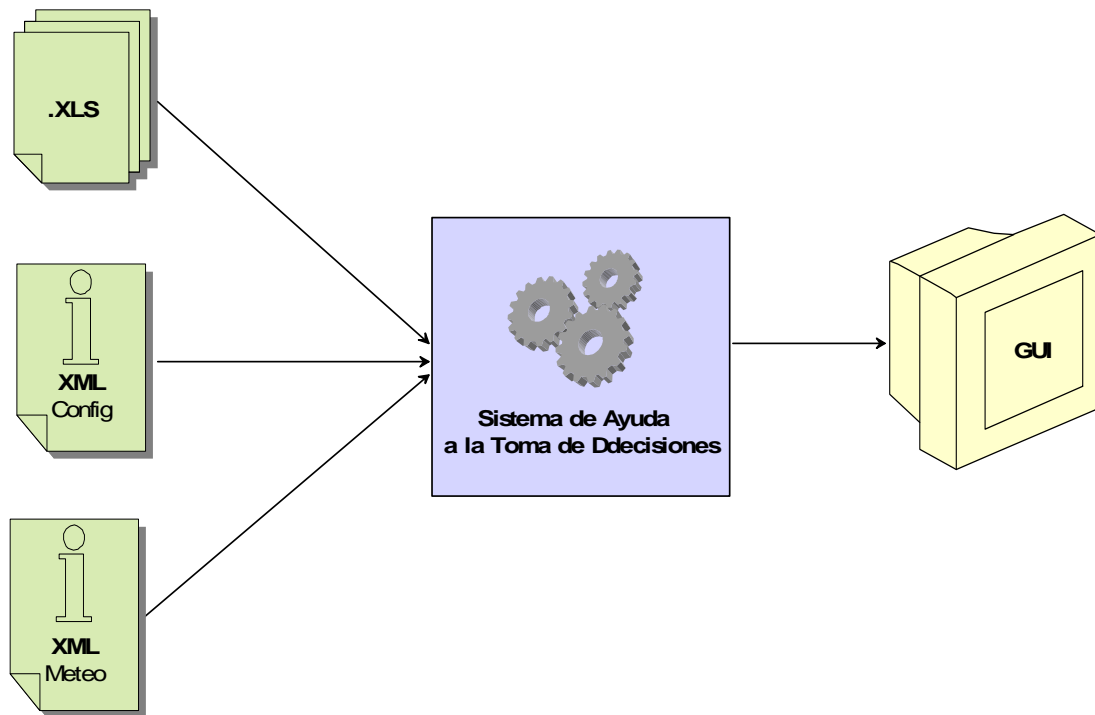


Figura 22: Flujo de datos en el SATD

El sistema diseñado recibe como entradas:

- un fichero Excel, donde se encuentran los datos numéricos suministrados por los sensores de la presa, que utilizará el SATD para realizar el razonamiento y la toma de decisión
- un fichero XML de configuración de parámetros de las funciones de pertenencia de las variables
- datos numéricos de variables meteorológicas, que facilitan entre otras cosas, el cálculo de la función que proporciona un valor a la variable Histórico. Se accede a ellas a través de un fichero XML facilitado por el Centro de Supercomputación de Galicia(CESGA), que el sistema descarga vía web.

La salida del sistema consiste en dos acciones fundamentales, por una parte se muestra al usuario, por pantalla:

- el resultado del razonamiento, esto es las decisiones tomadas, y
- el protocolo de actuación definido en los métodos de actuación de la presa.

Atendiendo a su funcionalidad, se puede distinguir cuatro módulos claramente diferenciados, Figura 23:

- Adquisición de datos (entradas)

- Preproceso de datos y Fuzzyficación de variables
- Razonamiento y toma de decisiones
- Visualización de resultados (salidas)

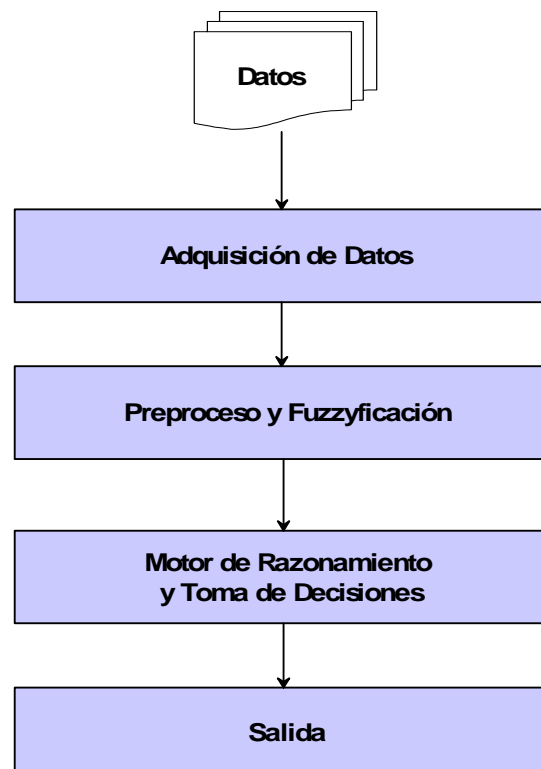


Figura 23: Diagrama de módulos del SATD atendiendo a su funcionalidad

### 7.5.1 Módulo de adquisición de datos

Este módulo gestiona la interacción del sistema con el exterior, en este apartado se profundiza sobre las entradas al SATD.

Para la adquisición de los datos en tiempo real, es necesario establecer un canal de comunicación con los servidores de datos de sensores o de recursos externos. Para ello se utilizará el paso de datos a través de un fichero Excel (.xls). En este fichero se almacenan los datos horarios referentes a la información sobre el estado de la presa, proporcionada por los sensores. El sistema obtiene el valor numérico de las variables que intervienen en el razonamiento, a partir de este fichero.

Con la idea de que este modelo de razonamiento pueda implantarse en otras presas con diferentes características a la propuesta en este escenario, se ha pretendido relajar las restricciones de definición de los parámetros estructurales de la presa, permitiendo que el usuario personalice la definición de las funciones de pertenencia de las variables implicadas. Esto se lleva a cabo mediante un fichero de configuración XML. En este

fichero, el usuario puede definir los parámetros que configuran las funciones de pertenencia de las variables que intervienen en el sistema, con el objetivo de ajustar al máximo las restricciones del proceso de razonamiento.

Un claro ejemplo de su utilidad puede verse sobre la variable **Cota**. Para el caso de la presa de Belesar, la etiqueta lingüística **Cota Alta**, se define a partir de unos valores que se ajustan a la altura del muro de contención de dicha presa. En el caso de presas de diferente tamaño, los valores que determinan la cuantificación lingüística de la altura de la cota serán diferentes, pues están estrechamente ligados y unívocamente determinados por los parámetros estructurales del embalse.

El fichero de configuración XML posee la siguiente estructura. Cada nodo del árbol XML corresponde a una variable. Éstos a su vez tienen tantos nodos hijos como posibles valores lingüísticos pueda tener la variable correspondiente. Los atributos de los nodos hijos corresponden a los cuatro puntos que definen la función trapezoidal de cada etiqueta lingüística.

```
<cota>
  <baja punto1 punto2 punto3 punto4>
  <media punto1 punto2 punto3 punto4>
  <alta punto1 punto2 punto3 punto4 >
</cota>
<aforo>
  <reducido punto1 punto2 punto3 punto4>
  <elevado punto1 punto2 punto3 punto4>
  <muy_elevado punto1 punto2 punto3 punto4
  >
</aforo>
<lluvia>
  <debil punto1 punto2 punto3 punto4>
  <moderada punto1 punto2 punto3 punto4>
  <fuerte punto1 punto2 punto3 punto4 >
  < muy_fuerte punto1 punto2 punto3
  punto4>
  <torrencial punto1 punto2 punto3 punto4>
</lluvia>
<gradienteCota>
  <lento punto1 punto2 punto3 punto4>
  <rapido punto1 punto2 punto3 punto4>
</gradienteCota>
<gradienteAforo>
  <lento punto1 punto2 punto3 punto4>
  <rapido punto1 punto2 punto3 punto4>
</gradienteAforo>
```

La tercera entrada del sistema es suministrada por el Centro de Supercomputación de Galicia (CESGA). Este organismo dispone de archivos XML con datos atmosféricos que adquiridos en múltiples estaciones meteorológicas. El sistema se conecta al servidor del centro y se descarga el archivo XML, a partir de su URL. Una vez descargado se extrae la información requerida. En este caso, la información que se obtiene del fichero, es por una



parte información redundante con la proporcionada por la presa, como puede ser la lluvia, mientras que por otra, aporta información relevante para la elaboración del valor de la variable Histórico. El valor denominado histórico es una función de promedios y desviaciones en períodos semejantes al actual en años anteriores, o en otro caso, agravantes o atenuantes de condiciones meteorológicas actuales.

El modulo de adquisición por tanto, se encarga de recopilar los datos procedentes de las diferentes fuentes para pasarlos al módulo de preproceso y fuzzyficación.

### 7.5.2 Módulo de preproceso y fuzzyficación

Este módulo recibe como entrada, la información procedente del módulo de recepción de datos. Dependiendo del carácter de la información, el tratamiento de la misma se realiza de dos formas diferentes:

1) En el caso de tratarse de los valores numéricos de las variables utilizadas en el proceso de razonamiento, se procede al cálculo del grado de pertenencia a la etiqueta lingüística de la variable correspondiente.

El modelo de representación de los valores de una variable fuzzy, se realiza mediante una función de pertenencia trapezoidal. Por tanto cada etiqueta lingüística de cada una de las variables, queda unívocamente determinada por cuatro puntos que describen los segmentos que definen la función, Figura 24.

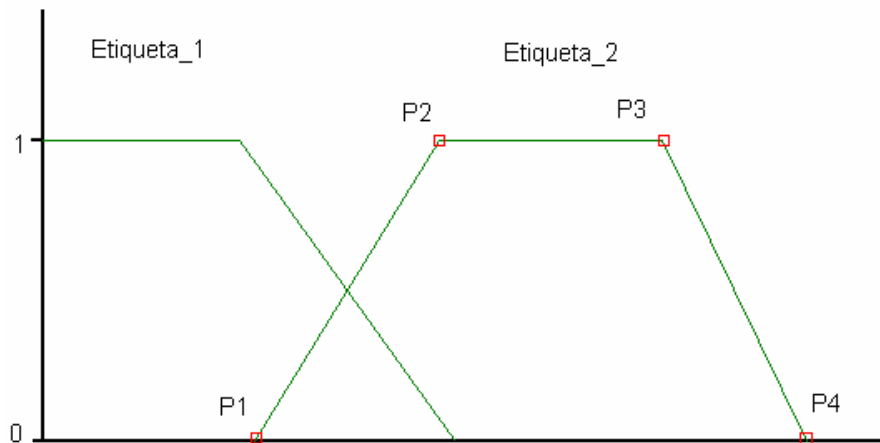


Figura 24: Representación gráfica de una variable fuzzy. Cada conjunto de cuatro puntos { P1, P2, P3, P4} determina la función de pertenencia de la etiqueta lingüística correspondiente.

Por tanto, el grado de pertenencia de un valor a una determinada etiqueta lingüística, viene determinado por la siguiente expresión:

$$\mu_{v^{etiq}}(x) = \begin{cases} ax + b, & x \in (-\infty, k] \\ c, & x \in [k, l] \\ dx + e, & x \in [l, \infty) \\ 0 \text{ e.o.c.} \end{cases}$$

Para el cálculo del valor lingüístico de una determinada variable, se ha aplicado la función máximo sobre los grados de pertenencia calculados para todas las etiquetas lingüísticas que componen la variable.

2) Por otra parte, en el caso de que la información provenga del fichero de configuración, se procede a la construcción de las funciones de pertenencia que determinarán los valores lingüísticos de las variables, en el caso de que se necesite ajustar dichas funciones para otro embalse. Para ello se obtiene del fichero de entrada, los cuatro puntos que definen la función de pertenencia trapezoidal de cada variable, requeridos para la definición de cada una de las funciones.

### 7.5.3 Módulo de Razonamiento y Toma de Decisiones

Este módulo recibe como entradas las variables lingüísticas previamente definidas y como salida, las alertas o indicaciones del sistema de razonamiento embebido en el mismo.

Una vez adquiridas y procesadas las variables que intervienen en el razonamiento, el sistema aplica el algoritmo de razonamiento aproximado, descrito en secciones anteriores.

El sistema recorre los nodos del árbol que corresponden a las variables, ramificando en función de la restricción establecida por el valor fuzzy de la variable, obtenido aplicando la función máximo sobre los grados de pertenencia a las etiquetas lingüísticas de la variable.

En cada nodo se conoce la información lingüística de la que dispone el sistema, por lo que únicamente se permite el acceso al conocimiento que verifica las restricciones previas, sobre los valores de las variables. De esta manera se accede única y exclusivamente a los bloques de la base de conocimiento que tienen relevancia para el estado actual.

### 7.5.4 Módulo de salida

En este módulo se muestran al usuario, las decisiones tomadas por el sistema de razonamiento con las actuaciones a realizar sobre las compuertas en relación al estado actual del sistema bajo vigilancia.

Como el sistema diseñado pensando en su aplicación a la presa de Belesar, la salida se ajusta a las condiciones estructurales de la misma. Esta presa dispone de 6 compuertas de vertido, sobre las que se puede actuar en grupos de 3, es decir, se pueden manejar 3 compuertas simultáneamente, o el grupo completo de 6. En cuanto a la apertura de las compuertas existen dos posibilidades, apertura media o apertura total. Por tanto, ordenando por riesgo creciente las salidas del sistema, tenemos:

- Compuertas cerradas
- 3 compuertas abiertas hasta la mitad
- 3 compuertas abiertas en su totalidad
- 6 compuertas abiertas hasta la mitad

- 6 compuertas abiertas en su totalidad

Debido a la naturaleza del dominio, la salida no puede inferirse por alguno de los métodos conocidos de defuzzyficación recogidos en [28],[29], ya que, la apertura o no de compuertas, está sujeta a unos protocolos de actuación que son determinados por el organismo correspondiente de la cuenca hidrológica a la que pertenece la presa. Por ello, el sistema diseñado también proporciona al usuario como salida adicional, el listado de los protocolos de actuación en la situación de riesgo detectada.

## 8 Resultados

En este apartado se muestran los resultados obtenidos tras las pruebas con el sistema de razonamiento aproximado que se ha diseñado.

Para realizar las pruebas del sistema se han utilizado datos proporcionados por el despacho de generación, de la central hidroeléctrica de Velle donde se encuentra el Centro de Control Integral (CCI). Los datos corresponden a un espacio temporal comprendido entre inicio del año 2000 y final del año 2001 en los que se alcanzaron diferentes niveles de riesgo: desde riesgo nulo, hasta riesgo máximo por avenida. Esto se debe a que la frecuencia de las avenidas tiende a seguir un ciclo repetitivo de 2 avenidas cada 7 años [30]. Son más interesantes para el estudio, los datos de esta época porque en la actualidad, no se ha producido ninguna situación significativa, en la que se puedan detectar diferentes niveles de riesgo.

El sistema dispone del conocimiento transferido por expertos en el dominio, mediante entrevistas. Por otra parte, del 100% de los datos facilitados, se ha analizado un 33% para refinar y matizar el conocimiento obtenido previamente por el experto. El resto se ha destinado a las pruebas del sistema. Para ello, se ha ejecutado el sistema utilizando como entrada el fichero con los datos que no han sido previamente analizados. De esta manera se intenta cubrir la mayor casuística posible, y enfrentar al sistema a nuevas situaciones. La diferencia de criterio con la decisión del experto, es la que marca la precisión y validez de la decisión tomada por el sistema. A continuación se muestra una comparativa de las decisiones tomadas por el experto y por el sistema.

Se ha decidido dividir los datos en grupos de 1000 elementos, para facilitar la visualización de los eventos. Las gráficas en color azul, Figuras 26.1 a 30.1, representan la decisión tomada por el experto, mientras que las gráficas en color rojo, Figuras 26.2 a 30.2, indican la salida del sistema para el mismo conjunto de datos.

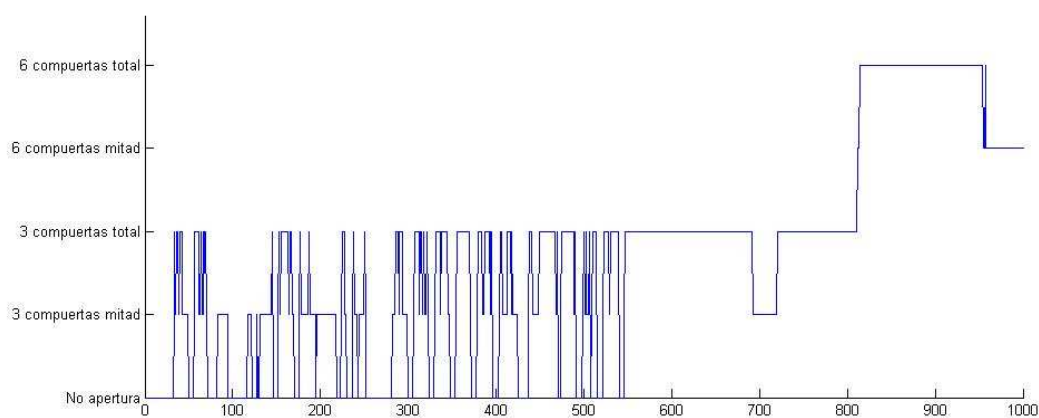


Figura 26.1 Decisión tomada por el experto, sobre la apertura de compuertas, para el primer grupo de 1000 datos

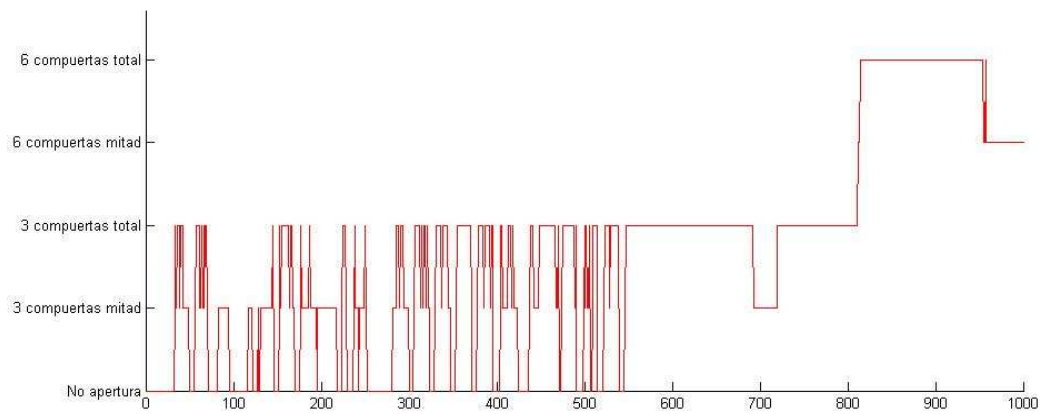


Figura 26.2 Decisión tomada por el SATD, sobre la apertura de las compuertas, para el primer grupo de 1000 datos

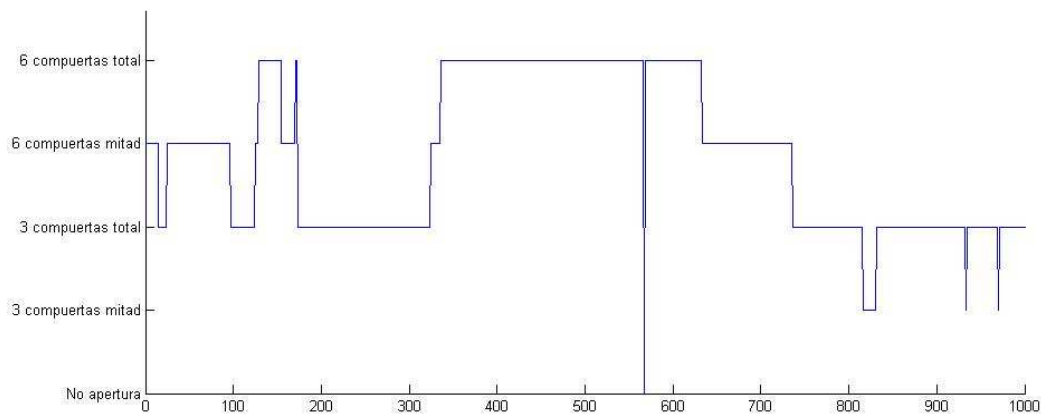


Figura 27.1 Decisión tomada por el experto, sobre la apertura de las compuertas, para el segundo grupo de 1000 datos

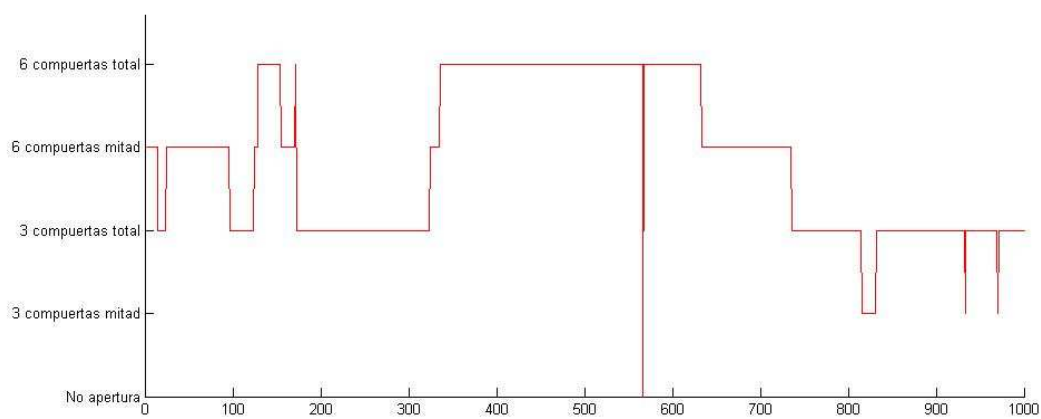


Figura 27.2 Decisión tomada por el SATD, sobre la apertura de compuertas, para el segundo grupo de 1000 datos

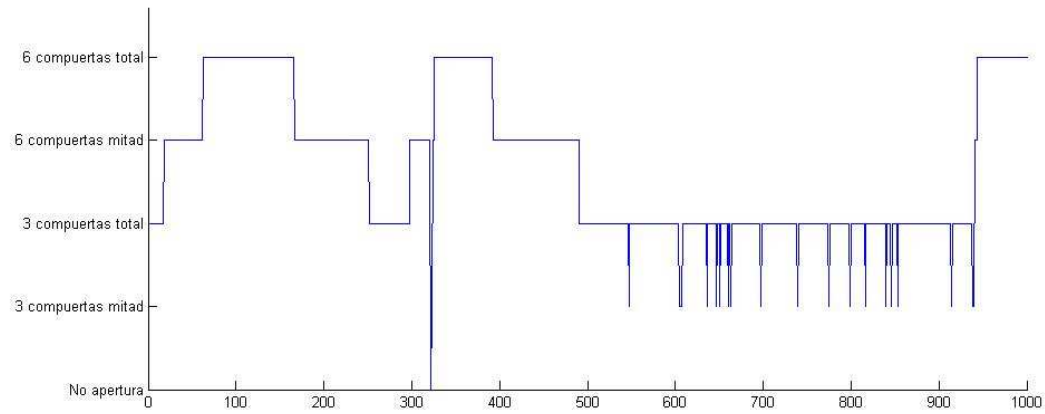


Figura 28.1 Decisión tomada por el experto, sobre la apertura de compuertas, para el tercer grupo de 1000 datos

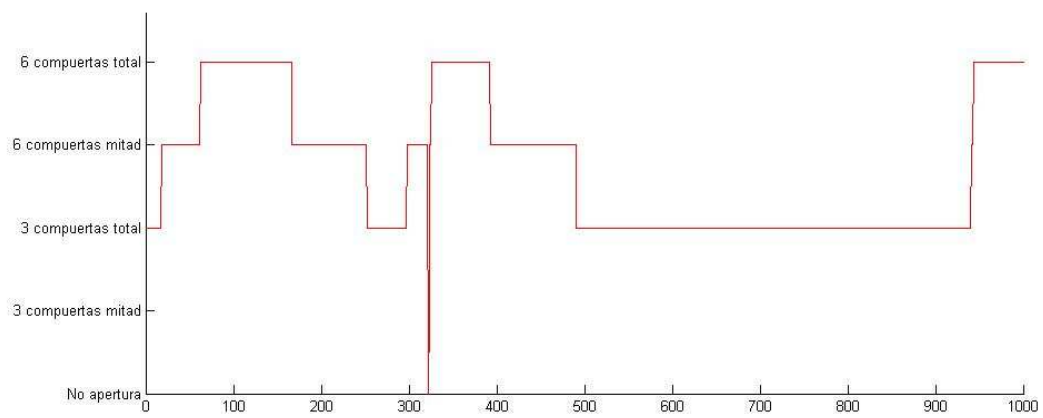


Figura 28.2 Decisión tomada por el SATD, sobre la apertura de compuertas, para el segundo grupo de 1000 datos

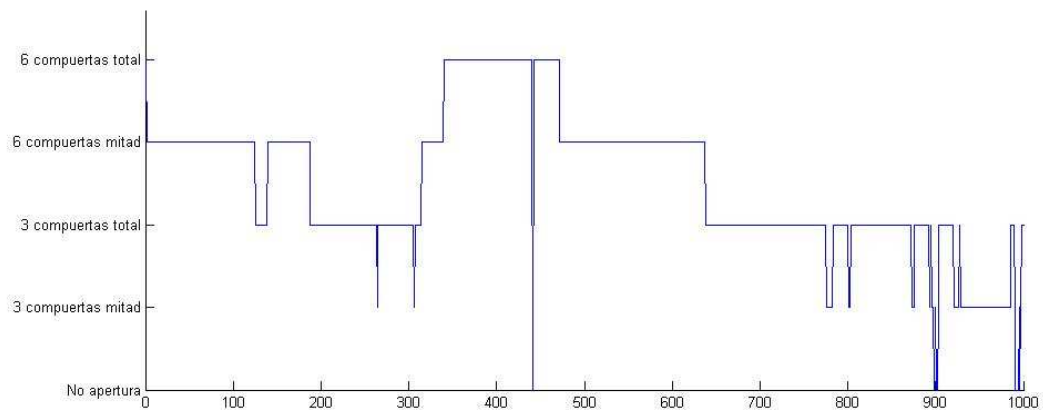


Figura 29.1: Decisión tomada por el experto, sobre la apertura de compuertas, para el cuarto grupo de 1000 datos

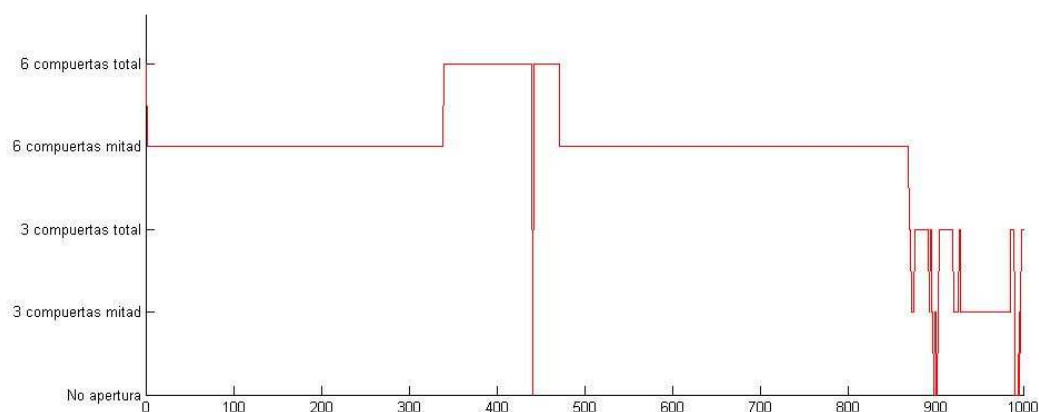


Figura 29.2: Decisión tomada por el SATD, sobre la apertura de compuertas, para el cuarto grupo de 1000 datos

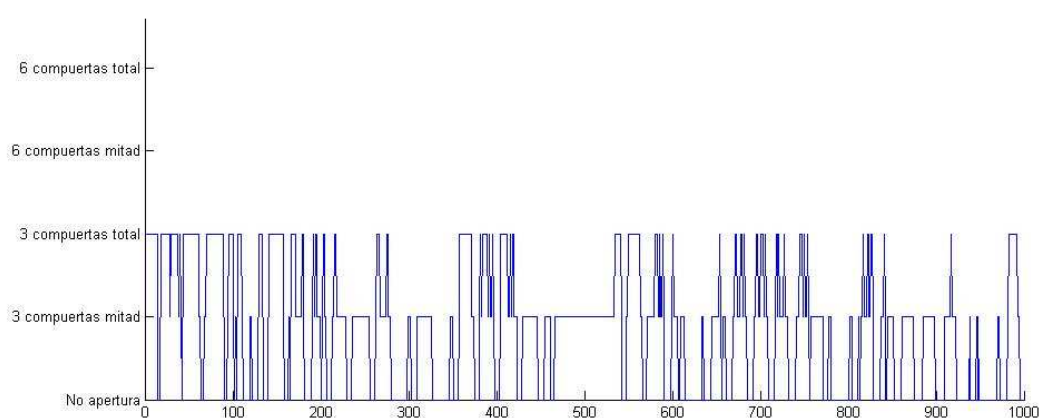


Figura 30.1: Decisión tomada por el SATD, sobre la apertura de compuertas, para el cuarto grupo de 1000 datos

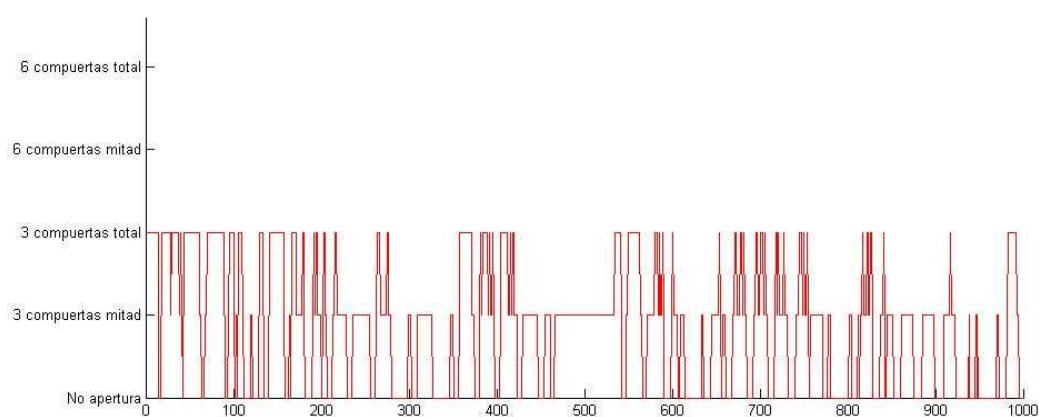


Figura 30.2: Decisión tomada por el SATD, sobre la apertura de compuertas, para el cuarto grupo de 1000 datos

Como se puede comprobar en las gráficas, las decisiones que toma el sistema de razonamiento coinciden en un alto porcentaje, en concreto, un 94.6%, con las tomadas por el experto. En las gráficas 26.1 con 26.2, 27.1 con 27.2, y 30.1 con 30.2, se observa que los resultados en cuanto a la decisión son idénticos. Cabe destacar las diferencias existentes

en las decisiones tomadas por el experto y el sistema automático, mostradas en las gráficas de las Figuras 28.1 y 28.2 y 29.1 y 29.2. En los grupos de datos en los que se presentan diferencias, estas se deben principalmente a situaciones anómalas, ya sea por presentar los datos valores inusuales, o porque las decisiones tomadas por el experto, no coinciden con la norma indicada en el protocolo y se desconoce que variables utilizó para variar la actuación.

En las Figuras 28.1 y 28.2 se puede observar que la diferencia en la decisión es muy leve. En estos casos, la causa radica en la falta de precisión de los datos medidos por los sensores. Una posible solución que evitaría esta diferencia de criterio consistiría en refinar la definición de los puntos que determinan las funciones de pertenencia, para intentar un mejor ajuste a las etiquetas lingüísticas que definen el valor fuzzy de las variables.

La diferencia de criterio que aparece en las gráficas de las Figuras 29.1 y 29.2 se debe a otro motivo. En este caso, según la Base de Conocimiento del sistema, y los datos recibidos, el sistema toma una decisión correcta y acorde a las circunstancias. El problema reside en que, en concreto para los períodos temporales en los que el sistema y el experto difieren, se dieron unas circunstancias excepcionales [26] que hicieron que el criterio del experto sobre la actuación de las compuertas cambiase con respecto al criterio de apertura del protocolo de actuación. Debido a situaciones como ésta, que pueden repetirse, aunque no dejan de ser excepcionales, surge la idea de incorporar al sistema la capacidad de aprendizaje. Esta tarea de investigación se realizará en trabajos futuros.

Salvo estas dos excepciones, la toma de decisión del sistema automático generado, se ajusta perfectamente al criterio marcado por el experto en la actuación sobre las compuertas de la presa, en el período analizado.



## 9 Comentarios, conclusiones y trabajo futuro

En este trabajo se presenta un Sistema de Ayuda a la Toma de Decisiones (SATD) capaz de razonar mediante el uso de árboles de decisión fuzzy, sobre el nivel de riesgo al que está expuesto un embalse o presa, y la consecuente apertura de compuertas, en función de determinadas variables de carácter tanto estructural como meteorológico.

El objetivo que se ha seguido, ha sido desarrollar un sistema automático que ayude a la vigilancia y control del nivel de agua, para su utilización en la presa de Belesar (Orense). Con este sistema de razonamiento aproximado, se intenta la detección temprana de situaciones pre-críticas y el envío de alertas.

Se ha propuesto como sistema de razonamiento un algoritmo de razonamiento que utiliza una representación en árboles de decisión fuzzy, que permite modelar con claridad los esquemas de razonamiento del operario en tareas de vigilancia y control. Así mismo, el uso de lógica fuzzy permite manejar convenientemente la incertidumbre y ambigüedad de las variables y conceptos relativos al dominio.

Se ha implementado un sistema buscando la flexibilidad de adaptación, a través de la opción de modificar los parámetros de definición de las funciones de pertenencia de las variables mediante un fichero de configuración XML. De esta manera, el sistema puede personalizarse para presas de diferentes características estructurales, con una única configuración inicial o incluso relajar o endurecer las restricciones sobre diferentes variables que afectan al volumen de agua del embalse, como puede ser la cota o el aforo. Se ha conseguido por tanto, un sistema de vigilancia, información y control, más completo y adaptable para determinar el estado de la presa.

La Base de Conocimiento ha sido verificada de acuerdo a los criterios utilizados por el experto, mediante entrevistas que tenían como finalidad, la adquisición de conocimiento sobre el dominio, variables que suelen intervenir en la toma de una decisión y ajuste de diferentes parámetros, tanto estructurales, como meteorológicos. Así mismo se ha refinado el conocimiento, para la toma de decisión, con una parte de los datos, de históricos, con valores de variables y acciones realizadas durante un determinado periodo de tiempo. La conjunción de las fuentes de conocimiento sobre el dominio, ha dado como resultado una Base de Conocimiento, que se ajusta con buenos resultados al conocimiento del operario experto.

La interfaz que se ha implementado permite mediante su visualización, la verificación de la Base de Conocimiento.

El demostrador implementado permite la verificación del sistema de reglas de acuerdo con los criterios utilizados por el experto, así como el ajuste de parámetros

Una posible ampliación en la que se está trabajando actualmente es en el diseño de un sistema capaz de generar motores de razonamiento gestores de crisis de distinta índole.

La idea consiste en abstraer el sistema de razonamiento de gestión de crisis basado en árboles de decisión fuzzy, independiente del contexto, es decir, extrapolar el modelo de razonamiento a otros dominios que presenten restricciones similares al problema (altamente no lineal, con un número elevado de variables y un grado alto de incertidumbre y ambigüedad en los conceptos) que se ha analizado.

Por un parte esta medida, al menos en una etapa inicial, debería contar con la colaboración de posibles usuarios finales. Estos usuarios deberían aportar al sistema información acerca de las variables que intervienen en el razonamiento, de su definición en términos lingüísticos, de la jerarquía de prioridades existente entre ellas y de sus interrelaciones. Para ello se plantean diferentes alternativas, entre ellas se encuentra la posibilidad de paso de toda esta información a través de un fichero de configuración que se cargue al iniciar el sistema. Otra posibilidad sería mediante una interfaz gráfica de configuración, más amigable para el usuario, en la que este podría introducir la información requerida para configurar el motor de razonamiento.

Otra de las posibles ampliaciones que puede admitir el sistema propuesto, que no excluye a la anterior, es la incorporación de aprendizaje. Se trataría de no limitar el sistema a mostrar el resultado del razonamiento como sugerencia a la acción a realizar, sino de tener en cuenta también una posible corrección del sistema por parte del usuario. Esto permitiría aprender de estas variaciones, para reajustar convenientemente la Base de Conocimiento. Para esta ampliación se han pensado diferentes métodos que soportan aprendizaje, como pueden ser las redes neuronales, o los sistemas basados en casos.

## 10 Referencias

- [1] J. M. de la Cruz, M. Santos, G. Pajares, J.J. Ruz, J. Quesada (2006), Evaluación de técnicas de IA para la toma de decisiones. *Nuevas Técnicas en Control Inteligente, Ed. Universidad de Huelva*, pp. 7-12.
- [2] Zadeh, L.A. (1965), Fuzzy sets, *Inform. Control*, 8, pp. 338-353.
- [3] Mitra, S., Konwar, K.M., Pal S.K. (2002) Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation. *IEEE Trans. Syst., Man, Cybern.*, vol. 32, pp. 1-12.
- [4] Albertos, P., Sala, A. (2004) El control borroso: una metodología integradora. *Revista Iberoamericana de Automática e Informática Industrial (RIAI)*, 1 (2) 22-31
- [5] Mitra, S., Konwar, K.M., Pal S.K. (2002) Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation. *IEEE Trans. Syst., Man, Cybern.*, vol. 32, pp. 1-12.
- [6] Janikow, C.Z. (1998) Fuzzy decision trees: Issues and methods. *IEEE Trans. Syst., Man, Cybern.*, vol. 28, pp. 1-14.
- [7] Siler, W. , Buckley, J.J (2005) Fuzzy Expert Systems and Fuzzy Reasoning. *Wiley Interscience*
- [8] C.S. Krishnamoorthy, Madras; S. Rajeev, Ramco (1996) Artificial Intelligence and Expert Systems for Engineers, Ed. *University of Hawaii, Honolulu, USA*, pp, 29-40.
- [9] Pino, R. , Gomez A., de Abajo N. (2001) Introducción a la inteligencia artificial: Sistemas expertos, redes neuronales y computación evolutiva, *Publicaciones Universidad de Oviedo* pp 10 – 12.
- [10] García-Alegre, M.C. (1991) Inteligencia Artificial en el control de procesos, controladores borrosos” *Mundo Electrónico*, 214, pp. 42-49.
- [11] Munakata T (2008), Fundamentals of the new Artificial Intelligence, *Springer-Verlag* pp.121 – 143.
- [12] Janikow, C.Z. (1998) Fuzzy decision trees: Issues and methods. *IEEE Trans. Syst., Man, Cybern.*, vol. 28, pp. 1-14.
- [13] Quinlan, J.R (1993), Programs for Machine Learning. *Morgan Kaufmann Publishers*.

- [14] Janikow C.J., Buiu C., Dumitrache I. (1994) "Learning fuzzy controllers by genetic algorithms," in *Proc. ACM Symp. Applied Computing*, pp. 232–236
- [15] R. Gallion, D. C. St. Clair, C. Sabharwal, and W. E. Bond, "Dynamic ID3: A symbolic learning algorithm for many-valued attribute domains," in *Proc. 1993 Symp. Applied Computing*. New York: ACM Press, 1993, pp. 14–20.
- [16] D. McNeill , P. Freiberger,(1993) *Fuzzy Logic*. New York: Simon and Schuster.
- [17] J.R. Quinlan (1986) "Induction on decision trees," *Mach. Learn.*, vol. 1, pp. 81–106.
- [18] J.R. Quinlan (1987), Decision trees at probabilistic classifiers, *Proc. 4th Internat. Workshop on Machine Learning* Morgan Kauffman, Los Altos, CA.
- [19] Yufei Yuan a\*, Michael J. Shaw b (1995) "Induction of fuzzy decision trees" *Fuzzy Sets and Systems* 69 pp. 125-139.
- [20] G.J. Klir (1987) "Where do we stand on measures of uncertainty, ambiguity, fuzziness and the like?" *Fuzzy Sets and Systems* 24 pp. 141 160.
- [21] Cristina Olaru, Louis Wehenkel, (2003) "A complete fuzzy decision tree technique" *Fuzzy Sets and Systems* 138 pp. 221–254.
- [22] A. Kandel and G. Langholz (1993), *Fuzzy Control Systems*, Eds. Boca Raton, FL: CRC..
- [23] Janikow C.J. (1996) A genetic algorithm method for optimizing the fuzzy component of a fuzzy decision tree," *Pattern Recognition*, Eds. Boca Raton, FL: CRC, pp. 253–282
- [24 ] [http://es.wikipedia.org/wiki/Presa\\_\(hidr%C3%A1ulica\)](http://es.wikipedia.org/wiki/Presa_(hidr%C3%A1ulica))
- [25] Colegio de Ingenieros de Caminos, Canales y Puertos de Madrid, (2003) Explotación ordinaria de la presa- embalse en relación al riesgo por avenida", *Jornada Técnica : Avenidas en Presas*.
- [26] Resumen reunión con personal CCI Presa Belesar (Orense – Mayo 2009).
- [27] Mandanero, S. (2000) Seguridad En Las Presas: Los Aliviaderos. [http://www.verticalia.com/revista/remortaje\\_hidro3.html](http://www.verticalia.com/revista/remortaje_hidro3.html)
- [28] E.H. Mamdani, J.J. Ostergaard, E. Lembessis (1983) "Use of Fuzzy Logic for implementing rule-based control of industrial process" *In: Advances of fuzzy sets, possibility theory and process* (Ed. P.P. Wang) Plenum Press

- [29] M.Sugeno, (1985) "An Introductory survey of fuzzy control" *Inf Sciences* 36, pp.59-83
- [30] Hesperia (2005), Caracterización de Escenario, Central Hidráulica de Belesar.

## 11 Apendice A: Diseño del SATD

En este apartado se presenta el proceso de diseño seguido para la implementación del Sistema de ayuda a la toma de decisiones. Se profundiza en detalles del diseño con orientación a objetos, sin llegar a la implementación, con el fin de mostrar en alto nivel la arquitectura del sistema, así como una leve descripción de la jerarquía de clases.

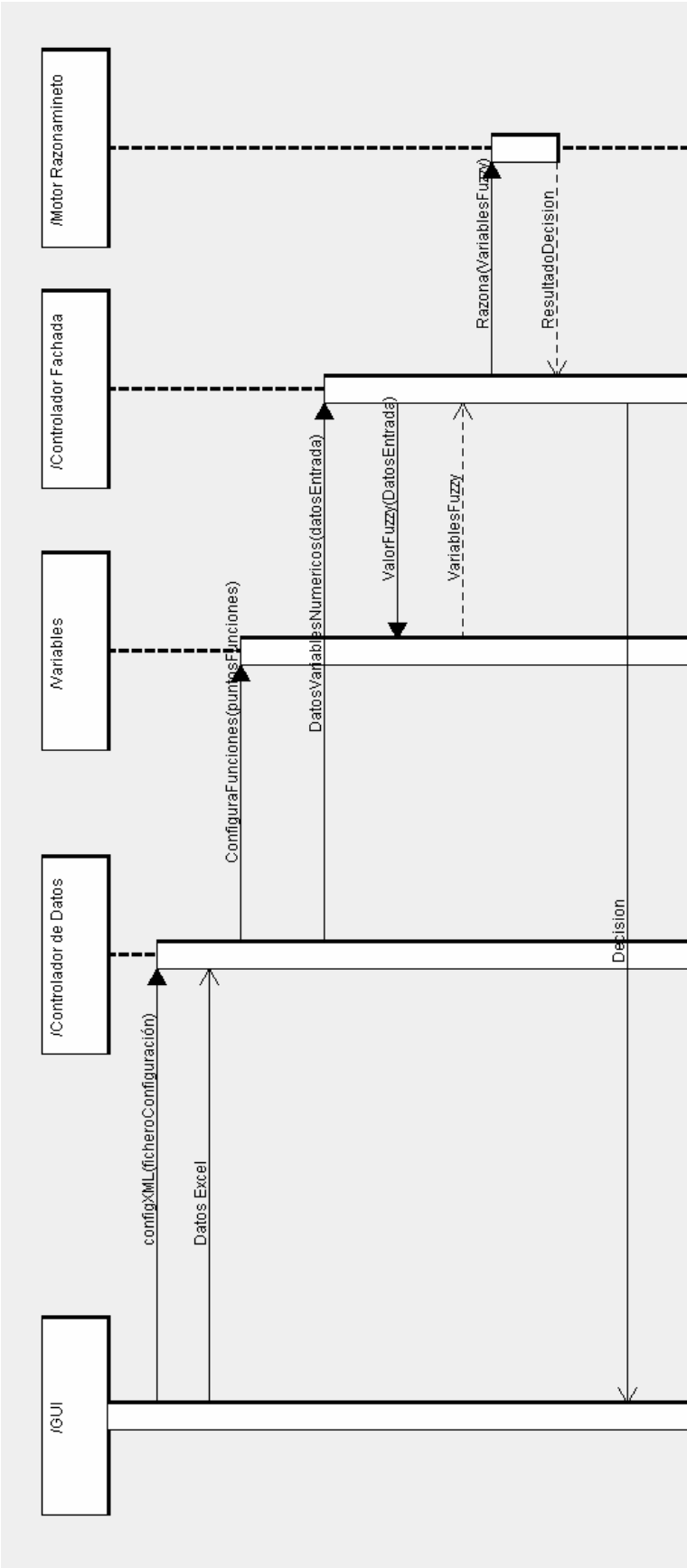
EL Sistema presenta un caso de uso principal, de acuerdo a la manera en que el usuario puede relacionarse con el sistema.

|  |
|--|
| <b>Descripción:</b><br>Aconsejar al usuario de la aplicación que hacer ante el estado actual de la presa y del medioambiente.  |
| <b>Actores:</b> Operario de la presa   |
| <b>Precondiciones:</b><br>Ninguna  |
| <b>Flujo Normal:</b><br><br><ol style="list-style-type: none"><li>1. El actor inicia el sistema por defecto</li><li>2. Si el actor no marca la casilla de obtención de datos Web ir a 4.</li><li>3. El sistema obtiene variables procedentes del CESGA</li><li>4. El sistema razona sobre el estado de las variables de la presa y del medioambiente.</li><li>5. Se muestra por pantalla, el nivel de riesgo y la actuación sobre las compuertas de la presa, y por otro lado el protocolo de actuación.</li></ol> |
| <b>Flujo Alternativo:</b><br><br><ol style="list-style-type: none"><li>2. El actor pulsa el botón de carga de fichero de configuraciones.</li><li>3. El sistema carga y configura de acuerdo con lo especificado en el fichero.</li></ol>  |
| <b>Postcondiciones:</b><br>El actor ha visualizado la decisión que sugiere el sistema de ayuda a la toma de decisiones.  |

A continuación, se muestran algunos diagramas UML con el fin de aclarar la arquitectura que se ha seguido en la implementación. Todas las clases que aparecen en los diagramas se corresponden a las principales clases del programa, el resto de ellas se han considerado detalles de implementación.

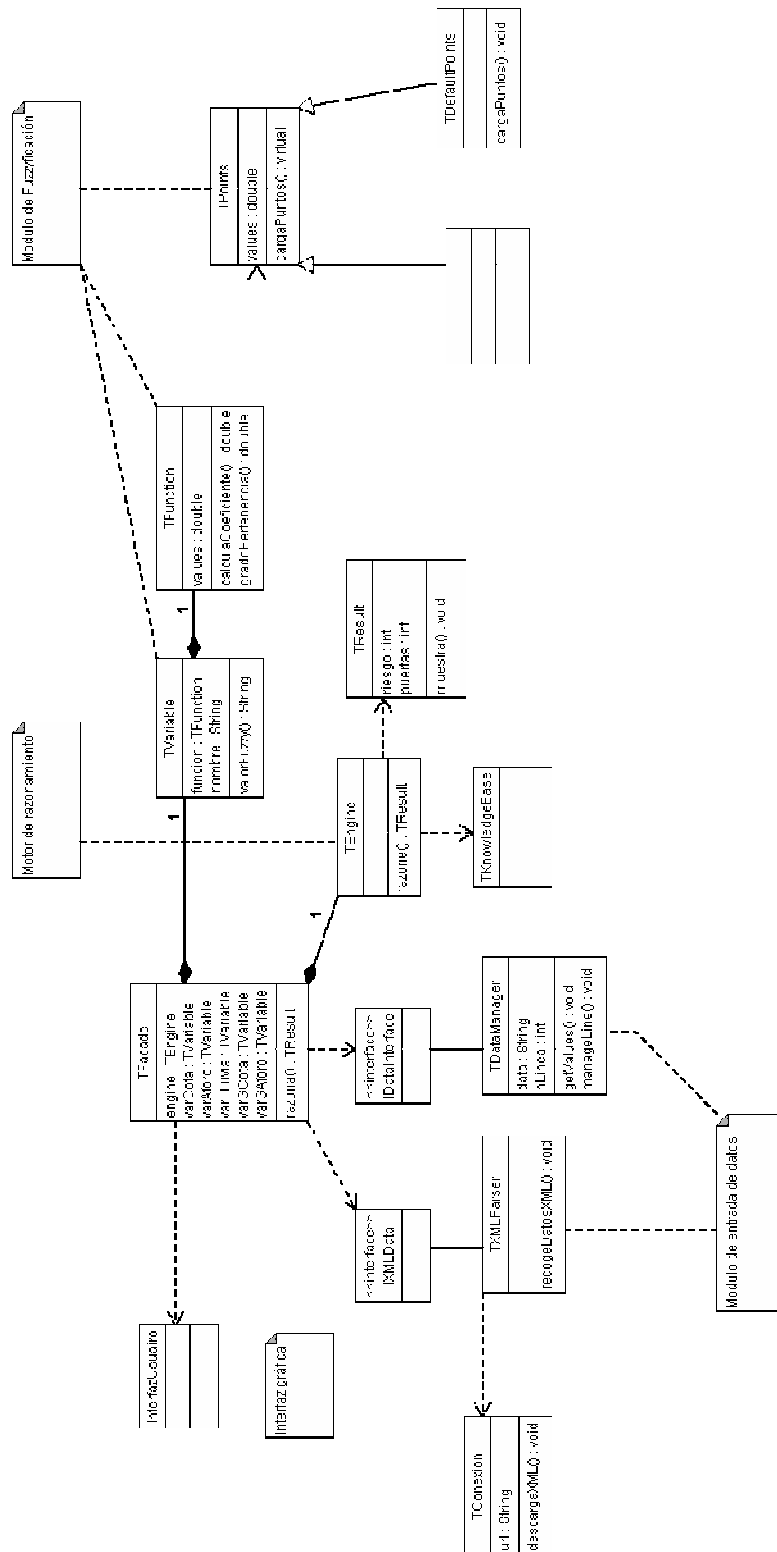
Para el diseño de la aplicación se han seguido directivas estándar de orientación a objetos, utilizando patrones de diseño.

Diagrama de secuencia.





### Diagrama de clases



## 12 Apéndice B: Implementación

En este apartado se explica con mayor detalle que en apartados anteriores la implementación de las clases principales del SATD. La sección se ha dividido en los principales paquetes del sistema, para mayor claridad y así contar con una visión global pero detallada de la estructura que se ha seguido en el proceso de implementación.

### 12.1 Paquete entrada

Este módulo cuenta con las siguientes clases principales:

- Interfaz IDataManage:

Esta interfaz se encarga de la relación con el fichero de datos de entrada, por una parte y por la otra se relaciona con la clase TFacade. La función de las clases que la implementan es, mantener una manera clara, modular y flexible de acceso a los datos. Sirve como intermediario entre la fuente de datos y el sistema.

Métodos:

- `getValues()`: Es el método principal, las clases que lo implementen deben asegurarse de proporcionar correctamente los valores de las variables contenidas en la fuente de datos, en este caso un fichero Excel

- Clase TDataManage:

Esta clase es la encargada de implementar la interfaz IDataManage.

Métodos:

- `getValues()`: Implementación del método de obtención de valores
- `manageLine()`: Método propio de la clase del que se vale `GetValues()`, que se encarga de parte del procesamiento de los datos de entrada.

- Interfaz IXMLData:

Esta interfaz actúa de comunicador entre los datos provenientes de un fichero XML y la lógica del programa.

Métodos:

- `getXMLData()`: Es el método principal, las clases que lo implementen deben asegurarse de proporcionar correctamente los

valores de las variables contenidas en la fuente de datos, en este caso un fichero XML

- Clase TXMLParser:

Esta clase implementa la interfaz IXMLData, y su función es la de extraer los datos procedentes de un fichero XML.

Métodos:

- getXMLData(): Implementación del método de obtención de datos

- Clase TConexion:

Esta clase se encarga de establecer la comunicación con el servidor web que proporciona el fichero XML. Para ello se ha utilizado la librería libURL. Se trata de una librería de transferencia de archivos usando sintaxis URL, libre y de fácil uso en el lado del cliente. Esta librería soporta subidas FTP, FTPS, HTTP, formularios de subida HTTP, certificados HTTPS, SCP, SFTP, TFTP, TELNET, DICT, FILE, LDAP, LDAPS, proxies, cookies, autenticaciones user + password, continuación de transferencia de ficheros y http proxy tunneling. LibURL es portable, y compila y trabaja de la misma manera, sobre diferentes plataformas incluyendo entre otras: Solaris, NetBSD, FreeBSD, OpenBSD, Darwin, HPUX, IRIX, AIX, Tru64, Linux, UnixWare, HURD, Windows, Amiga, OS/2, BeOs, Mac OS X, Ultrix, QNX, OpenVMS, RISC OS, Novell NetWare, DOS...

La clase TXMLParser utiliza una instancia de esta clase para realizar la conexión y descarga del fichero.

Métodos:

- downloadXML(): Este método se encarga de conectarse a una determinada URL y descargar el fichero correspondiente.

## 12.2 Paquete de Fuzzy

En este paquete se encuentran las clases TPuntos, TFuncion y TVariable. La principal función de este paquete es modelar a nivel computacional la lógica borrosa que utiliza el sistema.

Cada variable se representa mediante un objeto de la clase TVariable.

Para su representación, se ha optado por utilizar funciones de pertenencia trapezoidales, por su generalidad y por implicar menor complejidad en su representación computacional. Por otro lado, se ha supuesto que cada variable toma varios valores (términos del lenguaje natural) dependiendo de la variable. Por ello, cada objeto TVariable cuenta con varias instancias de la clase TFunción, una por cada posible valor que la variable puede tomar.

Cada objeto de la clase TFuncion, está unívocamente determinado por cuatro puntos que forman los tres segmentos que compone cada función trapezoidal. La clase estática y abstracta TDefaultPoints se encarga de almacenar dichos puntos, en el caso de aplicar los puntos por defecto. En el caso de que se decidan modificar las funciones de pertenencia para adaptar el sistema a otras condiciones, estos puntos se obtienen de la clase TPointsVar.

Modificando los atributos de estas clases, se pueden modificar las diferentes funciones de pertenencia, para modelar los valores de cada variable, dependiendo del dominio y del usuario.

El proceso de creación de una variable borrosa en el sistema, es el siguiente:

- o Se definen los puntos que van a determinar cada etiqueta lingüística
- o Se crean las funciones, añadiendo el nombre de cada etiqueta
- o Se añaden las funciones creadas a la variable.

Este módulo cuenta con las siguientes clases principales:

- Clase TPoints

Se trata de una clase abstracta de la que heredan dos clases, TDefaultPoints y TPointsVar. El usuario de la aplicación es el que decide qué forma han de tener las variables, antes de iniciar el proceso de razonamiento para la toma de una decisión. En función de esta elección, el sistema de ayuda a la toma de decisiones utiliza uno u otra fuente de datos para formar las variables. Para resolver esta situación, en la implementación se ha utilizado un diseño siguiendo el patrón Factory method

Métodos:

- loadPoints(): Este método es implementado por las clases hijas, y su función es la de obtener los puntos que definen cada etiqueta.

- Clase TDefaultPoints

Esta clase implementa los métodos de su clase padre TPoints, en este caso, es la encargada de cargar los puntos en el caso de que el usuario no decida modificar los valores por defecto.

Métodos:

- loadPoints(): Carga los puntos por defecto

- Clase TPointsVar

Esta clase implementa los métodos de su clase padre TPoints, en este caso, es la encargada de cargar los puntos en el caso de que el usuario decida cargar nuevos valores que definan los límites de las funciones de pertenencia.

Métodos:

- loadPoints(): Carga los puntos procedentes del fichero XML que aporta el usuario.

- Clase TFunction

Esta clase modela cada una de las funciones que componen la variable. Los atributos de la clase son, una lista de puntos que determina la forma de la función, y una cadena que indica el nombre de la etiqueta a la que corresponde.

Métodos:

- calculaCoeficiente(): Este método, calcula los coeficientes de la recta de cada segmento de la función, y los guarda en una lista. Por cada segmento se calculan los dos coeficientes a y b que corresponden a la ecuación de la recta  $Y = aX + b$ , es decir, por cada función el método devuelve 6 coeficientes.
- gradoPertenencia (): Este método calcula el grado de pertenencia de un determinado valor que se pasa por parámetro, a una de las funciones de una variable. La función está dividida en varios intervalos, dependiendo de la variable, que vienen determinados por los puntos mencionados anteriormente. El valor se compara con los límites de los intervalos para identificar a cual de ellos pertenece. En función de cuál sea su intervalo, se le aplicará la ecuación correspondiente. La ecuación se obtiene al haber calculado previamente los parámetros de cada una, con el método calculaCoeficiente. El resultado de aplicar la ecuación de la recta correspondiente, al valor que se le pasa por parámetro al método, es el grado de pertenencia de dicho valor a la etiqueta lingüística que representa la función.

- Clase TVariable

Esta clase modela el comportamiento de una variable borrosa (fuzzy). Los atributos de la clase son un vector con varios objetos de tipo TFuncion y una cadena, con el nombre de la variable. Cada objeto TFuncion representa una etiqueta lingüística que puede tener como valor una variable. Cada variable maneja diferentes valores lingüísticos, dependiendo de qué variable sea.

Métodos:

- `valorFuzzy()`: Este método calcula el valor borroso (fuzzy) de un determinado número que se le pasa por parámetro, y lo devuelve como una cadena, simulando una variable lingüística. El objetivo del algoritmo es obtener los diferentes grados de pertenencia procedentes de cada una de las funciones que componen la variable. El valor borroso (fuzzy) asociado al número será el de la función que presente mayor grado de pertenencia.

## 12.3 Paquete Razonamiento

Este paquete alberga todas las clases que únicamente intervienen en el razonamiento.

- Clase `TEngine`

Esta clase se encarga de razonar sobre las posibles alternativas para decidir la más conveniente, en función de los valores ya procesados y fuzzyificados.

Métodos:

- `Razona()`: Este método tiene como función razonar sobre las variables de entrada al sistema previamente tratadas por los módulos anteriormente descritos. Sobre él recae la máxima responsabilidad en la toma de la decisión. Para llevar a cabo la toma de la decisión, es en éste método en el que se aplica el algoritmo descrito en secciones anteriores de razonamiento con árbol de decisión fuzzy.

El resultado de la decisión es tratado con la clase `TResult`

- Clase `TKnowledgeBase`

Esta clase actúa como base de conocimiento, y alberga todo el conocimiento sobre el dominio del que dispone el sistema. Se trata de una clase estática y abstracta que contiene todas las reglas de inferencia.

## 12.4 Paquete de lógica de programa

En este paquete están el resto de clases que se encargan del control del flujo del programa, la principal es la siguiente:

- Clase `TFacade`

Esta clase se encarga de abstraer y relacionar el entramado de operaciones y cálculos de la lógica del sistema con las diferentes interfaces de relación con el usuario. Se ha seguido la idea de diseño del patrón Fachada (Facade).

Al iniciar el sistema, se crea una única instancia, que actúa como intermediaria y coordinadora del flujo del programa. El interfaz de usuario se relaciona con el objeto TFacade, al que envía los datos necesarios para la toma de la decisión, y es el objeto TFacade, el que se encarga de crear las instancias de objetos pertinentes para desencadenar la ejecución del algoritmo. Una vez finalizada la ejecución del algoritmo, se encarga de relacionarse con la interfaz de usuario nuevamente para enviar el resultado obtenido, para ser mostrado al usuario.

Métodos:

- Razona(): Este método recibe por parámetro, la información recogida por la interfaz de usuario. Tiene como función además, enviar las variables al paquete de fuzzyficación, para una vez creado el motor, se tomar la decisión correspondiente.

- Clase TResult:

Con esta clase se pretende modelar el resultado devuelto por el motor. Se trata de una clase con dos atributos, en uno de ellos se recoge el nivel de riesgo o emergencia estimado ante el estado actual de las variables del sistema, mientras que el otro atributo recoge la actuación sobre las compuertas

## 13 Apéndice C: Artículo presentado para Jornadas de Automática, Tarragona 2008

### SISTEMA DE RAZONAMIENTO BASADO EN ÁRBOL DE DECISIÓN BORROSO: ESTUDIO DE UN CASO

J. Rivas, D. Martín, M.C. García-Alegre  
Instituto de Automática Industrial  
Consejo Superior de Investigaciones Científicas  
28500 Arganda. Madrid {jrivas, dmartin, maria@iai.csic.es}

M. Santos  
Facultad de Informática  
Universidad Complutense. Madrid. {msantos@dacya.ucm.es}

#### Resumen

*En este trabajo se presenta un sistema de razonamiento cualitativo que utiliza como soporte un árbol de decisión borroso para la vigilancia del nivel de agua de los embalses y la generación de alertas tempranas en situaciones previas a un posible desbordamiento. El objetivo es detectar situaciones anormales para prevenir con antelación a la población circundante. La aplicación desarrollada tiene como entradas, valores relativos al estado del embalse, de los ríos que confluyen en el mismo y de las condiciones meteorológicas. El razonamiento experto se modela mediante un árbol de decisión borroso, cuyas variables son las etiquetas lingüísticas que maneja el operario humano en su tarea de vigilancia y control.*

**Palabras Clave:** Sistemas de razonamiento, árbol de decisión borroso, vigilancia.

#### 1 INTRODUCCIÓN

En el pensamiento científico tradicional la comprensión de un fenómeno depende de la capacidad de analizarlo en términos cuantitativos precisos. Sin embargo los esquemas de razonamiento humano sugieren mecanismos de inferencia aproximados en cuanto a la génesis de las unidades informáticas, conceptos, y al establecimiento de las relaciones entre ellos, reglas.

El cerebro aprovecha la tolerancia en precisión para codificar la información relevante mediante términos lingüístico que conllevan una relación aproximada con los datos cuantitativos, asociándolos a percepciones y generando acciones [4].

En este marco, Zadeh presenta las bases iniciales de la teoría de conjuntos borrosos [9] con gran auge hasta el momento actual, aplicándose a áreas tan variadas como son el control automático, la optimización, clasificación, reconocimiento de patrones, etc.

Por otro lado y dentro de los modelos de razonamiento, se encuentran los árboles de decisión caracterizados por su capacidad de dividir un proceso complejo de toma de decisiones, en una secuencia de decisiones simples encadenadas que proporcionan un esquema de razonamiento fácil de interpretar [8],[1].

La fusión de conjuntos borrosos (fuzzy) y árboles de decisión permite combinar el manejo de la incertidumbre y habilidades del razonamiento aproximado, con la sencillez de comprensión y fácil aplicación que ofrecen los árboles de decisión, potenciando su aplicabilidad en contextos complejos con incertidumbre [8], [3].

Los árboles de decisión fuzzy (FDT) asumen que todos los atributos del dominio o variables lingüísticas, se han definido previamente en forma de conjuntos borrosos [6]. Se diferencian de los árboles de decisión tradicionales en dos aspectos fundamentales:



- Utilizan un criterio de bifurcación basado en restricciones borrosas
- Los procedimientos de inferencia son diferentes.

La selección de los conjuntos borrosos que definen los términos lingüísticos que se usan para construir el árbol, se realiza en el algoritmo, siguiendo un determinado criterio [6]. Esto se debe a que en un modelo de representación mediante reglas borrosas, el valor numérico de una variable puede tener un grado de pertenencia distinto de cero, en varios conjuntos borrosos, y al no existir restricciones se consideran las contribuciones de cada una de las reglas activadas en el resultado final.

En este trabajo se presenta, un sistema de razonamiento soportado por árboles de decisión borrosa como ayuda a la vigilancia y control del nivel de agua en los embalses, con la intención de detectar situaciones pre-críticas y enviar alertas tempranas.

Se ha organizado siguiendo el siguiente esquema. En la sección 2 se presenta el escenario de la aplicación a tratar. Las variables del sistema de decisión se definen en el apartado 3. En la sección 4 se describe el diseño del sistema de decisión borroso. En la 5 la implementación del mismo, para terminar con las conclusiones.

## 2 DESCRIPCION DEL ESCENARIO

El escenario a analizar es un embalse genérico o presa, con aportes de uno o varios ríos. En situaciones de deshielo y lluvias el nivel del agua aumenta considerablemente con riesgo de superar el muro del dique de contención, Figura 1, [2].

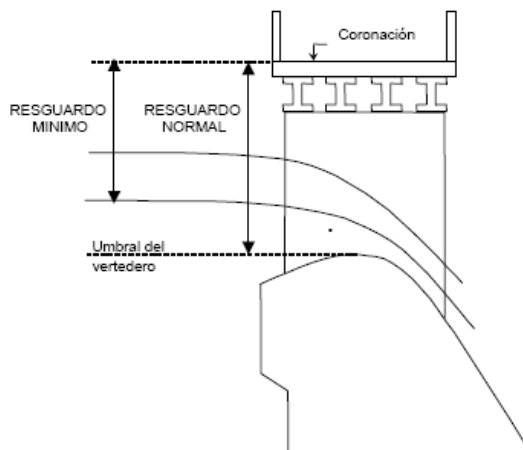


Figura 1: Sección transversal del muro de un embalse

A continuación se definen algunos términos del vocabulario del dominio, de interés para el diseño del sistema de razonamiento propuesto:

- Coronación: Superficie superior que delimita el embalse
- Aliviadero: Estructura hidráulica por la que rebosa el agua excedente cuando el embalse está lleno.
- Resguardo: Diferencia de cota entre el máximo nivel de agua del embalse o presa y el nivel de coronación
- Avenida: Elevación del nivel de un curso de agua muy superior a su flujo medio.

En un embalse se llega a una situación pre-crítica cuando el agua sobrepasa la coronación. Como consecuencia, pueden verse dañados elementos tales como las barandillas o el alumbrado e incluso, en situaciones extremas y dependiendo del tipo de construcción del muro de contención, podrían aparecer fisuras, por lo que es necesario prevenir estas situaciones mediante la detección del nivel de agua del embalse y las condiciones meteorológicas y medioambientales del momento de observación [7]

## 3 ANÁLISIS DE VARIABLES

El objetivo del sistema de razonamiento es doble:

- Vigilancia del estado del nivel de agua del embalse o presa
- Generación de alertas tempranas: riesgos previsibles, protocolos a activar y actuación sobre las compuertas en situaciones anómalas

La generación de alertas tempranas es fundamental para preservar tanto las instalaciones de la presa como las de la población asentada en las proximidades del mismo.

El primer paso en esta tarea consiste en realizar un análisis previo detallado de las variables que intervienen en el aumento del nivel del agua contenida en el embalse, pues la interrelación de las mismas va a determinar el nivel de alerta.

Las variables que afectan al nivel de agua de la presa son de índole atmosférico-medioambiental y tienen un fuerte grado de dependencia entre ellas, Figura 2. En el diseño del prototipo inicial de sistema de razonamiento, se han tenido en

cuenta tres variables: Lluvia, caudal del río y resguardo.

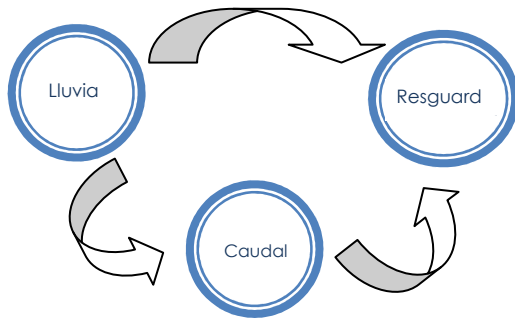


Figura 2: Diagrama de influencia de las tres variables seleccionadas. La punta de la flecha indica sobre qué variable se ejerce influencia

- **Lluvia:** Esta variable indica la cantidad de lluvia (mm/h) que afecta a la zona geográfica en la que se sitúa el embalse. La cantidad de lluvia incide tanto en el caudal del río como en el nivel del agua del embalse. Para que este valor sea realmente significativo es necesario tomar las medidas de lluvia en un punto lo más próximo posible al embalse.
- **Caudal:** Indica el flujo (en  $\text{m}^3/\text{s}$ ) del río que desemboca en el embalse. El caudal del río se ve influenciado por la lluvia, y afecta directamente al nivel de agua del embalse.
- **Resguardo:** Se define, en este dominio, como la diferencia en metros de cota entre el máximo nivel de agua embalsada en épocas de grandes crecidas del río y el nivel de coronación. Esto implica que cuanto menor sea su valor, mayor es el riesgo de vertidos por la coronación. Ésta es la variable más crítica, ya que está bajo la influencia de la lluvia y el caudal.

También se consideran otras variables que no dependen directamente del estado actual, pero se calculan a partir del modelo, experimental o histórico, de las variaciones del nivel del embalse, lluvia y caudal en años anteriores. La variable definida como histórico resulta de valorar la influencia de diferentes variables relacionadas con el contexto temporal, como son: la fecha actual del sistema, la estación del año, y otros condicionantes de riesgo.

El histórico permite comparar el estado de las variables en años anteriores con respecto al estado actual, para detectar posibles anomalías o variaciones que ayuden a detectar situaciones de pre-emergencia.

El tratamiento del histórico se contempla como una medida de prevención y contingencia de una situación de riesgo que no se ha dado todavía pero que, en caso de suceder, se desea evitar de inmediato.

## 4 DISEÑO DE LA APLICACION

Se ha diseñado una aplicación con interfaz gráfico para la visualización de estados y alertas, que se utilizará como demostrador y permitirá mostrar los resultados de la operación de diferentes algoritmos de razonamiento para la toma de decisiones. En este apartado se explica, en alto nivel, el diseño del programa, su estructura en bloques y el flujo de información.

Atendiendo a su funcionalidad, se puede distinguir cuatro módulos claramente diferenciados: Adquisición de datos, Preproceso de datos y Fuzzyficación, Razonamiento y toma de decisiones, y Salidas/Visualización de resultados, Figura 3.

Entradas  
Numéricas

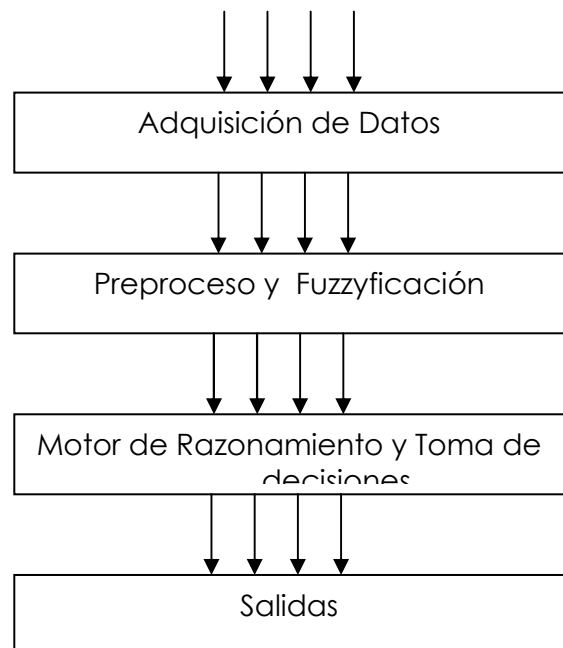


Figura 3: Diagrama de bloques del sistema

#### 4.1 MÓDULO DE ADQUISICIÓN DE DATOS

Este módulo gestiona la interacción del sistema con el exterior. Para la adquisición de los datos en tiempo real, es necesario establecer un canal de comunicación con los servidores de datos de sensores o recursos externos. Se encarga de la adquisición de todos los datos y del envío al módulo de preproceso y fuzzyficación. En concreto, en este primer prototipo, las entradas son:

- Nivel de Lluvia
- Caudal del río
- Nivel del resguardo del embalse
- Datos históricos

En el caso del histórico, el flujo de comunicación se establecerá con una base de datos, en la que estarán almacenados todos los valores de las distintas variables, referentes a años anteriores. El valor denominado histórico es una función de promedios y desviaciones en períodos semejantes al actual en años anteriores. El resultado es un valor lingüístico booleano: {Favorable, Desfavorable}.

#### 4.2 MÓDULO DE PREPROCESO Y FUZZYFICACIÓN

Este módulo recibe como entrada los valores numéricos y lingüísticos procedentes del módulo de recepción de datos. La salida del módulo son las mismas variables fuzzyficadas, es decir, el intervalo de posibles valores numéricos se engloba en un número reducido de valores lingüísticos, por ejemplo, {moderada, fuerte, muy fuerte}, Figura 4.

La computación con términos lingüísticos, requiere su representación mediante algún tipo de función o conjunto. En este caso se han representado mediante funciones trapezoidales, denominadas funciones de pertenencia, cada uno de los tres valores lingüísticos asignados a cada variable: {Lluvia, Caudal, Resguardo}.

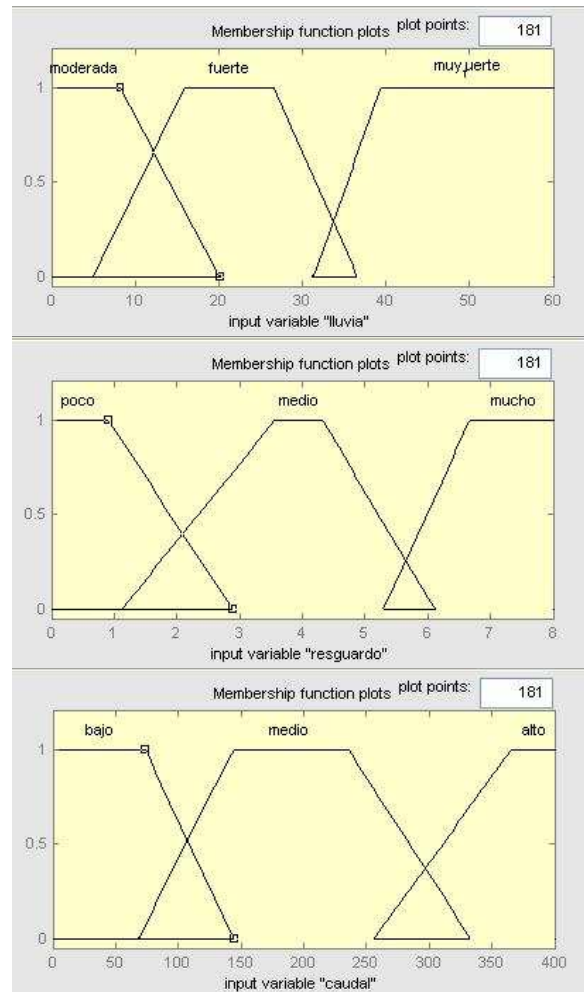


Figura 4: Representación gráfica de las funciones de pertenencia de las variables de entrada.

Para el caso de la lluvia los tres valores que puede tomar son: moderada, fuerte y muy fuerte; para el resguardo: poco, medio y mucho; y para el caudal: bajo, medio y alto.

#### 4.3 MOTOR DE RAZONAMIENTO Y TOMA DE DECISIONES

Este módulo recibe como entradas las variables lingüísticas previamente definidas y como salida, las alertas o indicaciones del sistema de razonamiento embebido en el mismo.

Las cuatro entradas son: las tres variables fuzzy, es decir, lluvia, caudal, resguardo, y la cuarta variable es un valor lógico (booleano) que resume el análisis del histórico.

Inicialmente se pensó en un sistema de reglas fuzzy, que integrasen el conocimiento experto y del dominio, del tipo:

**R<sub>i</sub>:** IF (entrada1 es vl<sub>1j</sub>) AND (entrada2 es vl<sub>2</sub>) AND (entrada3 es vl<sub>3j</sub>)

**THEN** (salida1 es vl<sub>1m</sub> AND salida2 es vl<sub>2n</sub>)

Donde la codificación de las variables es como sigue: vl<sub>ij</sub> es el valor lingüístico *j* asignado a la entrada *i*.

El planteamiento de un sistema de razonamiento borroso con una Base de Conocimiento borrosa no constituye la mejor solución en este caso por dos razones: 1) La combinación de las cuatro variables de entrada seleccionadas, tres de ellas con 3 valores lingüísticos, y una con dos, dificulta la generación de la Base de Conocimiento, y 2) La prioridad de unas variables sobre otras en los esquemas de razonamiento experto aconsejan otro modelo de razonamiento.

Las dos consideraciones anteriores, llevan a la propuesta de un modelo de toma de decisiones jerárquico representado mediante un árbol de decisión borroso [6], Figura 5.

En este árbol, cada variable puede tener mayor o menor prioridad en función de su relevancia en el dominio. La raíz del árbol es la variable con mayor prioridad, y de la que se parte, para discriminar el resto de valores de variables que no son útiles. En función de los valores lingüísticos de las variables se desciende por el árbol hasta llegar a las hojas terminales, dónde se encuentran las diferentes soluciones. A medida que se va descendiendo en niveles del árbol, se desbloquean conjuntos de reglas, por lo que solo se consultan aquellas que realmente tienen relevancia en el estado del sistema, favoreciendo el tiempo de cómputo. En el árbol, Figura 5, cada nodo simboliza una variable: Re resguardo, LL lluvia, C caudal, H datos históricos. Por otra parte en los nodos terminales hoja se representa el resultado del razonamiento: R nivel de riesgo y A indica la actuación de las compuertas.

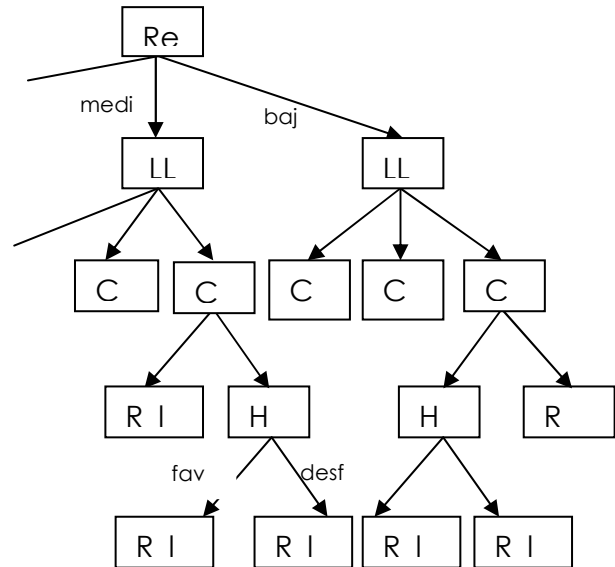


Figura 5: Extracto del árbol de decisión borroso.

Dado que no se requiere una salida numérica, en este dominio, no es necesario aplicar funciones de defuzzyficación. Los resultados de la etapa de razonamiento y toma de decisiones, que llegan al módulo de salida, son enviadas al interfaz del usuario.

Por defecto se ha considerado como raíz del árbol la variable resguardo de la presa (valor inverso al nivel de agua del embalse), por ser la más crítica. Esta asignación de prioridades queda en última instancia en manos del usuario, permitiendo la elección de la variable raíz en base a su propio criterio. De esta manera se flexibiliza la ponderación de las variables, posibilitando una mejor adaptación a las circunstancias de cada momento.

#### 4.4 MODULO DE SALIDA / INTERFAZ USUARIO

En este módulo se muestran al usuario, las decisiones tomadas por el sistema de razonamiento con las actuaciones a realizar, en relación al nivel de riesgo del estado actual del sistema vigilado: embalse y medioambiente, y a la necesidad o no de apertura de las compuertas de desagüe. También se deberá mostrar, en el interfaz de usuario, el conjunto de acciones a llevar a cabo atendiendo a los protocolos del sistema en situaciones riesgo tanto en pre-alerta como en aquellas que son críticas.

### 5. IMPLEMENTACIÓN DEL ARBOL DE DECISION BORROSO

Para la implementación del sistema de razonamiento representado mediante árboles de decisión borroso, se ha utilizado un diseño orientado a objetos, y para su implementación, lenguaje C++ , Figura 6.

Las principales clases con las que cuenta el sistema son:

- Clase TGui: Interfaz con el usuario
- Clase TFacade: Actúa como intermediaria entre la lógica de cálculo de la toma de decisión y la GUI. Su función es aislar la parte con mayor complejidad estructural del programa, de la interfaz de usuario (TGui). Se ha implementado siguiendo la estructura del patrón de diseño Facade [4].
- Clase TMotor: Se encarga de razonar con los valores borrosos (términos lingüísticos) de las variables. Se ha aplicado un patrón de diseño factoría simple, debido a que existe una clase correspondiente a cada tipo de motor en función de cada variable principal. Dado que el usuario decide en tiempo de ejecución cual va a ser la variable con mayor prioridad, se desconoce qué tipo de motor se va a crear.
- Clases TFuncion y TVariable : Estas clases tienen como finalidad modelar las variables fuzzy. Cada instancia de TVariable, está compuesta por tantas instancias de TFuncion como posibles valores lingüísticos pueda tener cada variable. En este caso se ha modelado cada variable con tres posibles valores.

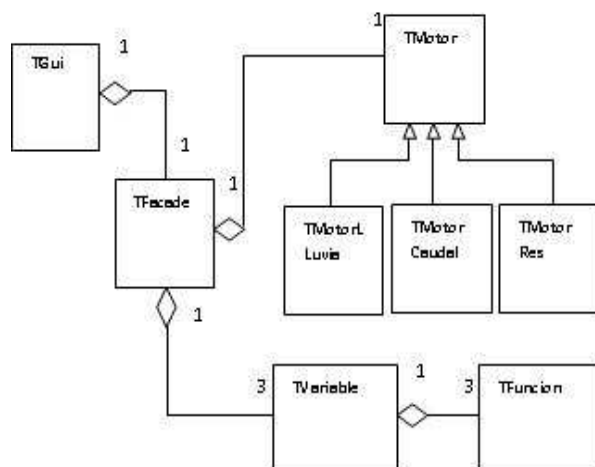


Figura 6: Diagrama UML de las principales clases del programa

Actualmente se dispone de una interfaz de desarrollador, como ventana al mundo del sistema de razonamiento sobre árbol de decisión fuzzy que permite mostrar la funcionalidad del sistema y verificar las decisiones propuestas con el usuario experto, Figura 7

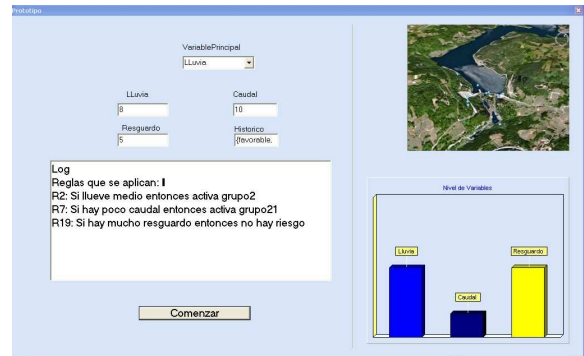


Figura 7: Captura de pantalla de la interfaz de desarrollador. En el menú de la izquierda aparecen el valor de las variables, y las reglas aplicadas.

## CONCLUSIONES

El sistema propuesto, basado en árbol de decisión borroso, permite modelar con claridad los esquemas de razonamiento del operario en tareas de vigilancia y control.

El demostrador implementado permite la verificación del sistema de reglas de acuerdo con los criterios utilizados por el experto, así como el ajuste de parámetros.

El sistema propuesto es fácilmente escalable tanto en el número de variables como en el de árboles necesarios para abordar cambios en las prioridades asignadas a las distintas variables.

